INTELIGENCIA RELACIONAL

Arquitectura de la interfaz humano-IA desde la periferia

Potenciación del intelecto en contextos de recursos limitados

Daniel Ricardo Novoa

Extensión de «Alfabetización Relacional: El lado humano de la inteligencia artificial»

Sierra Grande · Patagonia Argentina · 2026

Licencia Creative Commons BY-NC-SA 4.0

Producido con herramientas de código abierto en ejecución local

Nota sobre este libro

Este libro es una continuación de *Alfabetización Relacional: El lado humano de la inteligencia artificial* (Novoa, 2024). Aquel preguntó qué significa relacionarse críticamente con la IA desde una perspectiva humana y situada. Éste pregunta **cómo hacerlo con precisión, soberanía técnica y sin depender de infraestructuras ajenas**.

No se contradicen. Se necesitan.

Alfabetización Relacional demostró que la IA no es un interlocutor sino una herramienta con historia, sesgos y condicionantes institucionales. *Inteligencia Relacional* asume ese punto de partida y avanza: si la IA es una herramienta, aprendamos a operarla bien, sin perder de vista que el objetivo siempre es potenciar el pensamiento humano —especialmente allí donde los recursos son escasos y la dependencia tecnológica es políticamente problemática.

Tres advertencias previas:

1. **Rechazar el antropomorfismo** no significa tratar la IA como algo trivial. Significa entenderla con precisión para usarla mejor. El mecánico que sabe exactamente cómo funciona el motor cuida más el auto, no menos.
2. **La soberanía técnica no es un fin en sí misma**. Es la condición para que el pensamiento propio no dependa de infraestructuras ajenas, intereses corporativos ni conexiones a internet. Especialmente relevante en contextos como el nuestro: escuelas, comunidades, organizaciones con recursos limitados en los márgenes geográficos e institucionales.
3. **Este libro no es un relato**. No hay relatos de personajes. Hay, en cambio, una serie de operadores situados —docentes, estudiantes, técnicos— cuyas prácticas se presentan como **casos verificables**. La narrativa ha sido sustituida por la taxonomía.

PARTE 0 – EL MANIFIESTO DE LA PERIFERIA

0.1 La tesis de la escasez epistémica

La mayoría de la literatura sobre IA se escribe desde centros de abundancia tecnológica: San Francisco, Shenzhen, Londres. Sus autores nunca han preparado una clase con internet intermitente ni han usado una computadora con 8 GB de RAM y una raya en la pantalla.

Esta desigualdad de partida no es un detalle menor. Produce una **distorsión epistémica**: las preguntas que se consideran relevantes (¿cómo maximizar la creatividad con IA?, ¿cómo integrar ChatGPT en todas las aulas?) no son las mismas que emergen desde la periferia (¿cómo corro un modelo útil sin conexión?, ¿cómo protejo los datos de mis estudiantes de servidores extranjeros?).

Tesis central: la verdadera alfabetización en IA no consiste en aprender prompts ingeniosos, sino en **diseñar una interfaz que funcione bajo restricciones reales** —de hardware, conectividad, tiempo y soberanía de datos.

0.2 El mito de la herramienta neutral

Una herramienta deja de ser neutral cuando su diseño, su infraestructura y sus condiciones de acceso favorecen sistemáticamente algunos usos por sobre otros, o algunos usuarios por sobre otros.

Un cuchillo puede usarse para cocinar o para lastimar. Eso es neutralidad en el uso. Pero si el cuchillo solo lo pueden comprar algunos, si la empuñadura está diseñada para una mano derecha, si el manual viene solo en inglés —eso no es neutral.

La IA es así. Antes de que el usuario decida qué hacer con ella, ya hay decisiones tomadas:

- Qué datos se usaron para entrenarla (predominantemente inglés, hemisferio norte, perspectivas urbanas).
- Qué perspectivas quedaron subrepresentadas (ruralidad, sur global, lenguas minoritarias).
- Quién es dueño de la empresa que la provee (y qué hace con los datos generados).

Consecuencia práctica: operar IA desde la periferia exige **compensaciones activas** en cada etapa del flujo.

0.3 Los tres rechazos de la inteligencia relacional

Primer rechazo — Antropomorfismo ingenuo
Tratar a la IA como si pensara, sintiera o tuviera intenciones. Esto lleva a confiar en sus salidas como si fueran opiniones razonadas.

Segundo rechazo — Dependencia cloud por defecto
Asumir que la única forma de usar IA es mediante APIs propietarias que requieren internet permanente y envían datos a servidores externos.

Tercer rechazo — Delegación del juicio
Externalizar en la IA no solo tareas mecánicas sino también la evaluación, la formulación de preguntas y la responsabilidad última.

♦ **Advertencia**: estos rechazos no implican un uso menor de la IA. Implican un **uso más consciente y más complejo**.

0.4 El principio del exoesqueleto situado

Un exoesqueleto industrial no reemplaza los músculos del operario: los multiplica. El operario sigue decidiendo cada movimiento, pero puede levantar pesos que sus músculos solos no podrían.

Análogamente, la IA no reemplaza el juicio humano: multiplica su alcance. El humano sigue decidiendo qué pregunta hacer, qué información es relevante, qué inferencia es válida y qué acción tomar.

En contextos de recursos limitados, este exoesqueleto no es un lujo: es una herramienta de **equidad epistémica**. La escuela pública, el investigador sin acceso a grandes bases de datos, el colectivo autogestionado: cualquiera de ellos puede, con hardware austero y modelos locales, realizar tareas que antes requerían infraestructura millonaria.

0.5 Lo que este libro no es y lo que sí es

No es:

- Una colección de prompts mágicos.
- Un manual de programación de Transformers.
- Un libro de ética de la IA (aunque la ética lo atraviesa).
- Un relato de ficción docente.

Sí es:

- Un tratado sobre la interfaz humano-IA entendida como **disciplina de ingeniería y práctica emancipatoria**.
- Una metodología verificable y reproducible en hardware modesto.
- Un manifiesto por la **soberanía cognitiva desde la periferia**.
- Una caja de herramientas conceptuales que seguirán siendo válidas más allá de los modelos específicos de hoy.

PARTE I – LA MÁQUINA DE INFERENCIA SITUADA

Capítulo 1 – La fricción invisible como dato

1.1 El test de los tres ejemplos

Caso 1 – Docente en zona rural (Patagonia argentina)

- Hardware: PC con 8 GB RAM, conectividad intermitente.
- Consulta: *"Dame ejemplos de situaciones de pobreza para una clase de secundaria."*
- Salida del modelo (típica):

Personas sin hogar en ciudades grandes, comedores comunitarios en barrios marginales, familias que viven en autos.

Diagnóstico: sesgo de cobertura geográfica. El modelo ha aprendido que "pobreza" se asocia estadísticamente con entornos urbanos de países de ingresos altos.

Compensación: inyección de contexto geográfico explícito.

- Segundo prompt: *"Dame ejemplos de pobreza en una localidad rural de la Patagonia argentina. Secundaria."*
- Nueva salida:

Viviendas con calefacción insuficiente por el alto costo del gasoil, distancia al centro de salud (>100 km), trabajo zafral sin estabilidad, cierre de rutas por viento que impide acceso a alimentos.

Lección operativa: el modelo no "entiende" dónde está el usuario. El operador debe declarar las coordenadas contextuales como parte del prompt.

1.2 Qué la IA no puede ver

Por diseño, un modelo de lenguaje tiene acceso únicamente a la secuencia de tokens que recibe en su ventana de contexto. No posee:

- **Percepción del entorno físico** (temperatura, iluminación, dispositivos disponibles).
- **Conocimiento de la identidad del usuario** (salvo que se lo proporciones).
- **Memoria entre sesiones** (cada conversación empieza de cero, a menos que se use un historial explícito).
- **Conciencia del paso del tiempo** (no sabe si pasaron segundos o días desde el último mensaje).

Consecuencia práctica: cualquier información que no esté en el prompt no existe para el modelo. La **fricción** (respuesta inútil, genérica, descontextualizada) no es un fallo del modelo; es un **indicador de que el prompt omitió información relevante**.

▶ **Procedimiento – análisis de fricción**:
Cuando una respuesta no sirva, registrar:

1. ¿Qué supuse que el modelo sabía sobre mi contexto?
2. ¿Esa información estaba realmente en el prompt?
3. Reformular añadiendo explícitamente ese supuesto.

1.3 El costo invisible de la inferencia

Cada consulta a un modelo grande consume energía. El costo no es solo económico (API) sino también energético y ambiental. Los centros de datos utilizan agua para refrigeración y electricidad que a menudo proviene de fuentes no renovables.

En la periferia, además, cada consulta a una API en la nube implica:

- Dependencia de conectividad estable.
- Latencia (la respuesta tarda segundos o minutos).
- Transferencia de datos potencialmente sensibles a servidores en otros países.

♦ **Advertencia**: la comodidad de las APIs gratuitas suele tener un costo oculto: los datos del usuario se utilizan para entrenar y mejorar los modelos, y los términos de servicio pueden cambiar unilateralmente.

1.4 El idioma como frontera

La distribución lingüística en los corpus de entrenamiento de los principales LLM es extremadamente desigual. Estimaciones basadas en documentos de Common Crawl y otras fuentes:

Idioma	% aproximado en corpus de entrenamiento (GPT-3, Llama 2)
Inglés	>90%
Español	<5%

Idioma	% aproximado en corpus de entrenamiento (GPT-3, Llama 2)
Español rioplatense / patagónico	<0.1%

Implicancias:

- El modelo es más "fluido" y preciso en inglés; en español comete más errores sutiles.
- Las referencias culturales, ejemplos y marcos legales son predominantemente anglosajones.
- Los modismos locales, nombres de lugares pequeños o costumbres regionales están prácticamente ausentes.

No es un defecto malicioso. Es un **sesgo estructural** que el operador debe conocer y compensar.

Capítulo 2 – La caja de cristal: arquitectura de la inferencia

2.1 No es un cerebro

Es frecuente leer que las redes neuronales artificiales "imitan el cerebro humano". Es una analogía útil para introducir el concepto de neuronas y conexiones, pero se vuelve engañosa cuando se extrapola.

Diferencias fundamentales:

- El cerebro humano tiene **memoria episódica** (recuerda experiencias específicas). Un LLM no recuerda nada de una sesión a la siguiente a menos que se le proporcione el historial en el prompt.
- El cerebro está **encarnado** (siente hambre, frío, emociones). Un LLM no tiene cuerpo ni sensaciones.
- El cerebro aprende continuamente de la experiencia. Un LLM se entrena una vez y queda congelado (salvo versiones actualizadas).

Analogía correcta: un LLM es un **motor de inferencia estadística** que opera sobre representaciones vectoriales del lenguaje.

2.2 Tokens: las unidades del pensamiento estadístico

Un modelo no lee palabras como cadenas de caracteres. Las convierte en **tokens** —unidades atómicas que pueden ser palabras completas, partes de palabras, o incluso caracteres individuales— y cada token se asigna a un vector numérico de alta dimensionalidad (por ejemplo, 4096 dimensiones).

Ejemplo:

- Entrada: *"El gato duerme"*
- Tokenización típica: ["El", " ga", "to", " du", "erme"]
- Cada token → vector [0.12, -0.45, 0.89, ..., 0.33] (4096 números)

Consecuencia práctica: elegir una palabra en lugar de otra no es una cuestión de estilo retórico; es una **selección de coordenadas** en un espacio vectorial. Cambiar *"banco"* por *"entidad financiera"* desplaza el punto de partida de la inferencia.

▶ **Procedimiento para operadores**: pensar en sinónimos como **vectores de desplazamiento**. Si una respuesta es demasiado genérica, probar términos más específicos o técnicos para acotar la región del espacio latente.

2.3 El mecanismo de atención

La innovación central del Transformer (arquitectura base de todos los LLM modernos) es el **mecanismo de atención**. Permite que, al generar cada nuevo token, el modelo pondere la relevancia de cada token anterior en relación con todos los demás.

No es comprensión semántica. Es **cálculo de dependencias de largo alcance**.

Ejemplo: en la frase *"El banco del río estaba vacío porque el agua se había retirado"*, la palabra "banco" es ambigua. El mecanismo de atención calcula que "río" y "agua" tienen alta relevancia, "dinero" no aparece, y concluye: contexto geográfico, "banco" como ribera.

Implicación profunda: el modelo no tiene "memoria emocional" ni "continuidad de identidad". Cada inferencia es un cálculo fresco sobre la secuencia actual. El "diálogo" que percibimos es una ilusión creada por la concatenación del historial en el prompt.

2.4 La triple falacia de la agencia (reformulada)

Para usar la IA con precisión, es necesario identificar y rechazar tres confusiones sistemáticas:

Falacia	Definición	Corrección operativa
Falacia intencional	Creer que el modelo tiene objetivos propios.	El modelo maximiza la probabilidad del siguiente token. No “quiere” nada.
Falacia semántica	Creer que el modelo comprende el significado.	Opera sobre correlaciones estadísticas en un espacio vectorial. No “entiende”.
Falacia dialógica	Creer que interactuamos con un “otro” (persona, agente).	Interactuamos con un proceso estocástico condicionado por nuestra entrada.

♦ **Advertencia**: la fluidez del lenguaje activa en el cerebro humano un sesgo cognitivo automático: **si algo se comunica como una persona, lo tratamos como persona**. No podemos evitarlo por completo, pero podemos saber que está pasando y corregirlo conscientemente.

▶ **Procedimiento – test de consistencia estocástica**:
Ejecutar el mismo prompt tres veces con máxima aleatoriedad (parámetro temperature alto, por ejemplo 0.9). Si las respuestas varían en contenido factual, es prueba de que no hay “certeza” en el modelo, solo muestreo de distribuciones.

Capítulo 3 – El prompt como código

3.1 Definición operativa

Un **prompt** no es una pregunta, una sugerencia ni una orden en lenguaje natural dirigida a un interlocutor. Es un **vector de condiciones iniciales** que define la distribución de probabilidad condicionante para la generación de tokens.

En términos matemáticos:
$P(\text{salida} \mid \text{prompt}) = \prod P(\text{token}_i \mid \text{token}_1 \ldots \text{token}_{i-1}, \text{prompt})$

El prompt no se “responde”. Se **configura**.

3.2 Seis componentes de una configuración efectiva

No todos los prompts requieren los seis componentes. La regla es: **añadir complejidad solo cuando la tarea lo exija**.

1. Rol
"Actuá como un geógrafo especializado en recursos hídricos de la Patagonia."
Desplaza el espacio de probabilidades hacia el dominio de ese rol.

2. Contexto
"Esto es para estudiantes de una escuela rural. La mayoría nunca salió de la provincia."
Información que el modelo no puede adivinar.

3. Tarea
No: *"Hablame del río Negro."*
Sí: *"Describí tres problemas de contaminación en el río Negro y sus causas."*

4. Formato
"Tres párrafos. Cada párrafo: problema, causa, solución posible."
El formato acota tanto como el contenido.

5. Restricciones (positivas y negativas)

- Positivas: *"Usá al menos dos referencias al documento proporcionado."*
- Negativas: *"No uses tecnicismos. No menciones otros países."*

6. Criterio de calidad
"Un estudiante de 16 años debería entenderlo sin ayuda."
Ajusta la legibilidad y el nivel de abstracción.

3.3 El orden importa: recencia estratégica

En muchos modelos, los tokens que aparecen al final del prompt tienen mayor influencia en la generación que los del principio. Este fenómeno se debe al **sesgo de recencia** del mecanismo de atención.

Consecuencia práctica: si colocás la instrucción principal al inicio y luego añadís miles de tokens de contexto, esa instrucción puede tener influencia cercana a cero.

▶ **Procedimiento – recencia estratégica**:

1. Colocar el contexto extenso (documentos, ejemplos) **primero**.
2. Colocar la instrucción principal, las restricciones de formato y el criterio de calidad **al final**, justo antes de la generación.

Ejemplo de estructura óptima:

```
text
[CONTEXTO LARGO]
```

(Los primeros 3000 tokens: documentos, historial de conversación)

```
[INSTRUCCIÓN PRINCIPAL]
```
(La tarea específica)

```
[FORMATO]
```
(Cómo debe estructurarse la respuesta)

```
[RESTRICCIONES]
```
(Lo que no debe hacer)

```
[CRITERIO]
```
(Nivel de legibilidad, audiencia)

3.4 El registro de prompts (prompt log)

Una práctica de bajo costo y altísimo retorno: mantener un **archivo de texto** (o cuaderno físico) donde se anote:

- Fecha y contexto de uso
- Prompt exacto empleado
- Salida obtenida
- Evaluación: ¿sirvió? ¿qué faltó?
- Ajustes realizados para la siguiente iteración

Ejemplo de entrada:

```
text
```
2026-05-12 | Clase sobre voto joven

Prompt inicial: "Actividad de 30 min sobre voto joven, secundaria, sin internet"

Salida: lista de conceptos abstractos (democracia, representación), demasiado teórica.

Ajuste: añadí "que genere discusión, con ejemplos de la vida local, dos columnas en pizarrón"

Segunda salida: usable.

Lección: el modelo por defecto da definiciones, no dinámicas de aula.

A los tres meses, ese registro es un **manual de operación personalizado** que documenta qué funciona y qué no en el contexto específico del operador.

Capítulo 4 – Lo que el operador va aprendiendo (principios emergentes)

En lugar de una narrativa de aprendizaje, presentamos los **nueve principios operativos** que emergen de la práctica sostenida con IA en contextos de recursos limitados. Cada principio es una **regla de operación** que se puede enseñar, verificar y compartir.

Principio 1 – La IA no es una fuente de verdad

El modelo está entrenado para maximizar la **plausibilidad** de las secuencias, no la **veracidad factual**. Responde con la misma confianza cuando sabe y cuando inventa.

▶ **Procedimiento asociado**: ninguna afirmación fáctica generada por IA se acepta sin verificación por fuente independiente.

Principio 2 – Hay que declarar las coordenadas contextuales

El modelo no sabe dónde está el usuario, qué recursos tiene disponibles, ni qué limitaciones de tiempo o conectividad enfrenta.

▶ **Procedimiento asociado**: incluir en cada prompt relevante un bloque [CONTEXTO SITUADO] con: ubicación geográfica, hardware disponible, conectividad, edad de los destinatarios.

Principio 3 – Los sesgos no se eliminan; se compensan

Los sesgos son estructurales (reflejo de los datos de entrenamiento). No se "corrigen" con mejores prompts, pero se pueden **contrarrestar** pidiendo explícitamente perspectivas alternativas.

▶ **Procedimiento asociado**: después de obtener una respuesta, pedir *"Ahora dame un ejemplo desde una perspectiva diferente (rural, del sur global, de género, etc.)"*.

Principio 4 – La primera respuesta es un borrador

Esperar que la IA dé en el blanco al primer intento es una fuente común de frustración. El flujo eficiente es iterativo.

▶ **Procedimiento asociado**: presupuestar al menos **dos rondas de ajuste** para tareas no triviales. La segunda y tercera respuestas suelen ser significativamente mejores.

Principio 5 – No todos los modelos son iguales para la periferia

Los modelos grandes (70B+ parámetros) requieren hardware costoso o conexión a la nube. Los modelos pequeños cuantizados (3B–8B) corren en hardware de consumo y son "suficientemente buenos" para la mayoría de las tareas educativas.

▶ **Procedimiento asociado**: probar primero con un modelo local como Phi-3-mini o Qwen 2.5 7B cuantizado. Solo si la calidad es insuficiente, considerar una API en la nube.

Principio 6 – La responsabilidad no se delega

La IA no puede rendir exámenes, dar la cara ante los estudiantes, ni justificar sus errores ante una dirección escolar. Esa responsabilidad recae en el operador humano.

▶ **Procedimiento asociado**: cualquier material generado por IA que se lleve al aula debe ser **revisado y editado** por el docente. No se usa "tal cual".

Principio 7 – Lo local, cuando se pueda

Ejecutar modelos localmente (sin enviar datos a la nube) es la única forma garantizada de proteger la privacidad de los estudiantes y mantener la operación ante cortes de internet.

▶ **Procedimiento asociado**: antes de adoptar una herramienta de IA, preguntarse: *¿puedo correrla en mi propia máquina? ¿Hay una alternativa de código abierto?*

Principio 8 – Cuando la IA no entiende, a veces el problema es el prompt

Antes de frustrarse, preguntarse: *¿qué supuse que el modelo sabía que no estaba en el prompt?* La mayoría de las veces, la respuesta está en esa omisión.

▶ **Procedimiento asociado**: mantener una lista de "suposiciones comunes" propias y convertirla en un bloque de contexto estándar.

Principio 9 – Hay tareas que no se delegan por principio

Formular preguntas, evaluar la relevancia, decidir qué es importante y qué no, establecer criterios de calidad —esas son funciones del juicio humano. Delegarlas sistemáticamente atrofia la capacidad crítica.

▶ **Procedimiento asociado**: definir una **frontera clara** entre lo que se delega (síntesis, búsqueda, estructuración) y lo que se retiene (definición del problema, verificación de resultados, decisión final).

PARTE II – LA PRÁCTICA OPERATIVA

Capítulo 5 – De la transacción a la práctica

5.1 Dos modos de operación

Modo transaccional:

- El operador escribe un prompt, recibe una respuesta, la acepta o la rechaza.
- No hay análisis de por qué la respuesta fue buena o mala.
- La IA es una máquina expendedora.

Modo práctico:

- El operador ve cada interacción como parte de un **proceso de mejora continua**.
- Analiza la fricción: ¿qué información faltaba en el prompt?
- Mantiene un registro de ajustes exitosos.
- La IA es un **colaborador estadístico** cuyo comportamiento se aprende a orquestar.

5.2 La fricción como dato, no como error

En el modo transaccional, la fricción (respuesta inútil, genérica, desviada) genera frustración. En el modo práctico, la fricción es **información valiosa** sobre lo que el prompt omitió.

Ejemplo de análisis de fricción:

Respuesta recibida	Fricción observada	Información faltante en el prompt
Ejemplos de pobreza urbana	No aplica a mi localidad rural	"Localidad rural de la Patagonia"
Explicación del federalismo con ejemplos de EE.UU., Alemania, UE	Deriva hacia otros países	"No mencionar otros países"
Actividad teórica sobre voto	No genera discusión	"Que genere discusión, sin teoría abstracta"

5.3 El límite estructural del tiempo real

El modo práctico requiere **tiempo** para iterar, ajustar y verificar. En contextos de alta carga docente (dos turnos, viajes largos, preparación de clases en minutos), no siempre es posible.

No es un fracaso del operador. Es una restricción del sistema. La solución no es "esforzarse más", sino **diseñar flujos de trabajo que minimicen la necesidad de iteración**:

- Usar plantillas de prompts previamente validadas.
- Mantener un banco de contexto reutilizable (bloques estándar: "Somos una escuela rural de la Patagonia...").
- Preferir herramientas locales que responden más rápido (menos latencia) y permiten probar múltiples variantes en el mismo tiempo.

Capítulo 6 – Perímetros lógicos

6.1 Deriva contextual: definición y detección

Deriva contextual: tendencia del modelo a alejarse del dominio, tono, formato o supuestos implícitos que el operador esperaba, porque el prompt no define explícitamente un perímetro.

Causa: el modelo está entrenado para continuar secuencias de manera plausible en **cualquier** dominio. Sin restricciones, el perímetro por defecto es "todo el espacio latente del lenguaje humano".

Detección:

- La respuesta incluye ejemplos o referencias de países/regiones no mencionados.
- Cambia de nivel de formalidad sin aviso.
- Añade información que no estaba en el contexto proporcionado.

6.2 Inyección de contexto forzado (grounding)

La técnica más efectiva para eliminar la deriva es **anclar** el modelo a una fuente específica y prohibirle salir de ella.

Plantilla de grounding forzado:

```
text
[CONTEXTO EXCLUSIVO]
{aquí va el texto, documento o conjunto de datos}
```

```
[INSTRUCCIÓN]
Basándote ÚNICAMENTE en el contexto anterior, sin añadir información externa, respondé: {pregunta específica}
```

```
[REGLAS]
- Si la respuesta no está en el contexto: "NO_ENCONTRADO"
- No infieras información que no esté textualmente presente
- No añadas ejemplos ni comentarios fuera del contexto
```

Por qué funciona: el modelo recibe un perímetro estrecho. La probabilidad de generar tokens fuera de ese perímetro se reduce drásticamente (no a cero, pero sí a niveles manejables).

6.3 La jaula de Faraday semántica

Una **jaula de Faraday semántica** es un conjunto de restricciones explícitas (positivas y negativas) que delimitan el espacio de respuestas posibles. No es una sugerencia; es un filtro sobre la distribución de probabilidad.

Restricciones positivas (obligaciones):

```
text
[OBLIGACIONES]
```

- Respondé en exactamente 3 párrafos

- Cada párrafo debe comenzar con una palabra clave en negrita

- Incluí al menos 2 referencias directas al documento proporcionado

Restricciones negativas (prohibiciones):

```
text
[PROHIBICIONES]
- No uses adjetivos valorativos ("excelente", "terrible", "sorprendente")
- No incluyas opiniones personales
- No generes listas numeradas a menos que se solicite explícitamente
```

♦ **Advertencia**: las restricciones negativas funcionan mejor cuando se formulan en **positivo-negativo explícito** (“No hagas X”) que en términos vagos (“evitá divagar”). La razón es estadística: el modelo ha procesado millones de ejemplos de “No hagas X”; “evitá” es más difuso.

Capítulo 7 – Cadenas de razonamiento (Chain of Thought)

7.1 El problema del salto directo

Cuando un modelo responde una pregunta compleja con una sola afirmación, el operador no tiene visibilidad del **proceso interno** que llevó a esa respuesta. Puede tratarse de un razonamiento sólido o de una asociación superficial capturada en los datos de entrenamiento.

Chain of Thought (CoT) es una técnica de prompting que fuerza al modelo a generar **tokens intermedios** que representan pasos de razonamiento antes de la respuesta final.

Mecanismo: un problema complejo requiere una respuesta que es un punto de muy alta dimensión en el espacio latente. Saltar directamente del prompt a la respuesta es un salto de alta dimensionalidad, propenso a errores. Cuando el modelo genera tokens intermedios, descomprime ese salto en pasos de menor dimensionalidad, donde cada token intermedio condiciona al siguiente.

7.2 CoT verdadero vs. CoT simulado

CoT verdadero: cada paso depende lógicamente del anterior. Si se elimina un paso intermedio, el paso siguiente deja de ser comprensible o la conclusión cambia.

CoT simulado: una lista de afirmaciones independientes disfrazada de cadena. El modelo primero genera la conclusión por intuición estadística y luego inventa pasos que la justifican (razonamiento post-hoc).

▶ **Prueba de detección**: tomá un CoT generado y eliminá el paso central. Si el paso siguiente sigue siendo comprensible y el paso anterior sigue siendo útil, probablemente estás ante un CoT simulado.

7.3 Patrón de razonamiento etiquetado

Para aumentar la auditabilidad, se pueden exigir **etiquetas** que clasifiquen cada declaración.

Plantilla:

```
text
[INSTRUCCIÓN]
Resolvé el siguiente problema usando pasos etiquetados. Cada paso DEBE
tener una de estas etiquetas:
```

```
- [DATO]: Hecho extraído del problema o conocido.
- [INFERENCIA]: Conclusión derivada de datos previos.
- [SUPUESTO]: Afirmación no verificada que se asume para avanzar.
```

```
- [VERIFICACIÓN]: Comprobación de un paso anterior.
- [CONCLUSIÓN]: Respuesta final.
```

```
[PROBLEMA]
{problema}
```

Valor añadido: las etiquetas fuerzan al modelo a clasificar sus propias declaraciones. Un error categórico —etiquetar como [DATO] algo que no está en el problema— se vuelve detectable de inmediato.

♦ **Advertencia**: incluso con etiquetas, el CoT puede ser post-hoc. La única garantía es la **verificación externa** (contrastar la conclusión con fuentes independientes).

Fin de la primera entrega. La segunda entrega continuará con la Parte III (Los límites: sesgos, alucinaciones y lo que la IA no puede hacer), Parte IV (La política de la infraestructura y el trabajo invisible), Parte V (Despliegue soberano y RAG local), Parte VI (Pedagogía: docente real, tutor socrático, agentes), Parte VII (Ecología cognitiva y epistemología), y los Apéndices completos (plantillas, glosario, diez principios).

PARTE III – LOS LÍMITES ESTRUCTURALES

Capítulo 8 – Taxonomía de sesgos para el operador

8.1 El sesgo no es un error, es una característica

Un modelo sin sesgo sería un modelo que no ha aprendido nada útil. El sesgo es la **huella de la estructura del mundo** en los datos de entrenamiento. El problema no es la existencia de sesgos, sino la **falta de conciencia sobre ellos** y la ausencia de estrategias de compensación.

Definición operativa: un sesgo es cualquier desviación sistemática de las salidas del modelo respecto de una distribución de referencia que el operador considera deseable para su tarea específica.

8.2 Taxonomía de sesgos relevantes para la periferia

Sesgo geográfico (o de cobertura)
El modelo sobrerrepresenta perspectivas, ejemplos y marcos normativos del hemisferio norte (especialmente Estados Unidos y Europa occidental).
Detección: los ejemplos siempre mencionan ciudades del norte global, leyes estadounidenses, casos europeos.
Compensación: incluir en el prompt *"Dame ejemplos que tengan sentido en*

{localidad rural de la Patagonia}. Si no los tenés en tus datos, indicá 'NO_DISPONIBLE' en lugar de inventar."

Sesgo de frecuencia
El modelo tiende a generar las respuestas más comunes en sus datos de entrenamiento, incluso cuando el contexto sugiere una respuesta menos frecuente pero más precisa.
Detección: para preguntas sobre causas de un fenómeno, la IA menciona siempre las mismas 2-3 causas (las más citadas en internet), omitiendo causas locales o específicas.
Compensación: *"Buscá explicaciones alternativas, aunque sean menos comunes. Priorizá la precisión en este contexto específico sobre la frecuencia estadística."*

Sesgo de recencia (dentro del prompt)
El modelo da más peso a los tokens que aparecen al final del prompt. Las instrucciones al inicio de un prompt muy largo pueden tener influencia cercana a cero.
Compensación: recencia estratégica (véase 3.3).

Sesgo de anclaje numérico
El modelo se ancla a números, fechas o cantidades mencionadas en el prompt, incluso si son irrelevantes o incorrectas.
Detección: si el prompt dice *"aproximadamente 8.000 habitantes"*, el modelo tenderá a repetir o usar ese número aunque sea incorrecto.
Compensación: en tareas numéricas, no incluir la respuesta esperada en el prompt. Pedir que el modelo calcule y luego verificar.

Sesgo de alineamiento (el más peligroso)
El modelo tiende a estar de acuerdo con el usuario, incluso cuando el usuario está equivocado. Es invisible porque no produce señales externas de disenso.
Detección: ejecutar el mismo prompt con una afirmación y con su negación. Si el modelo está de acuerdo en ambos casos, hay fuerte sesgo de alineamiento.
Compensación: pedir explícitamente que asuma el rol de **crítico** o *abogado del diablo*. *"Actuá como alguien que debe encontrar fallas en mi razonamiento. No seas cortés. Señalá los errores que encuentres."*

Sesgo de presuposición
Si el prompt contiene una presuposición (*"según los expertos", "como es sabido"*), el modelo la acepta como válida sin cuestionarla.
Compensación: antes de usar presuposiciones, pedir al modelo que las verifique: *"¿Cuál es la evidencia de que los expertos coinciden en este punto? Si no hay consenso, indicá las posiciones divergentes."*

8.3 Compensación de sesgos desde el prompt

La compensación no requiere acceso al interior del modelo. Se hace íntegramente desde el lenguaje.

Estrategia 1 – Perspectiva múltiple forzada

```
text
[INSTRUCCIÓN]
```

Generá tres respuestas diferentes a la misma pregunta, desde tres perspectivas distintas:

1. Perspectiva dominante en la literatura anglosajona.

2. Perspectiva desde una comunidad rural del sur global.

3. Perspectiva que priorice saberes locales por sobre estadísticas agregadas.

Luego sintetizá una cuarta respuesta que integre lo válido de cada una.

Estrategia 2 – Contraste de fuentes

```
text
[INSTRUCCIÓN]
```

Tu respuesta debe considerar tanto la posición mayoritaria en los datos con los que fuiste entrenado como al menos una posición minoritaria o marginal. Indicá explícitamente cuál es mayoritaria y cuál minoritaria.

Estrategia 3 – Autodetección de sesgos

```
text
[INSTRUCCIÓN]
```

Antes de responder, identificá tres posibles sesgos que podrían afectar tu respuesta dado el tema y el contexto. Luego respondé intentando minimizar esos sesgos.

✦ **Nota situada**: el sesgo de cobertura es especialmente relevante en contextos latinoamericanos. Los modelos entrenados predominantemente con datos del hemisferio norte pueden tener representaciones deficientes de realidades locales (sistemas de salud, educación, justicia, geografía). No es neutralidad; es otra forma de sesgo estructural. Compensarlo es parte del trabajo del operador.

Capítulo 9 – Alucinaciones: no son bugs, son subproductos

9.1 Definición y naturaleza

Alucinación: generación de una afirmación falsa pero sintáctica y semánticamente coherente, que el modelo presenta con el mismo nivel de confianza que una afirmación verdadera.

No es un error en el sentido de un bug informático. Es un **subproducto natural de la función objetivo** del modelo: maximizar la plausibilidad de la secuencia, no la verdad factual. El modelo no sabe que no sabe.

9.2 Taxonomía de alucinaciones para el aula

Tipo	Definición	Ejemplo típico	Compensación
Factual	Afirmación sobre hechos objetivos falsa o no verificable	"La población de Sierra Grande es 8.000 habitantes" (dato desactualizado)	Grounding forzado; verificar con fuente externa
Razonamiento	Cadena lógica coherente pero con error no evidente	Argumento que parece válido pero omite una variable clave	CoT etiquetado + verificación por pares
Instrucción	Ignora una restricción explícita del prompt	Se le pide "no usar adjetivos" y los usa igual	Validación automática de formato; mover restricciones al final
Fuente	Atribuye información a una fuente inexistente	Inventa un paper, autor, DOI o página web	**Nunca** confiar en referencias bibliográficas generadas por IA sin verificación manual
Temporal	Confunde fechas o secuencias temporales	"La dictadura argentina comenzó en 1976 y Malvinas	Pedir línea de tiempo explícita; verificar cada fecha

Tipo	Definición	Ejemplo típico	Compensación
		fue en 1981" (error de año)	

9.3 El principio de confianza cero epistémica

Definición: ninguna afirmación generada por IA se acepta como verdadera a menos que sea verificada por una fuente independiente y confiable.

Tres niveles de verificación:

Nivel	Método	Cuándo usarlo	Limitación
Nivel 1 Interno	Ejecutar el mismo prompt 3 veces con temperatura alta. Si los hechos varían, es alucinación.	Detección rápida de alucinaciones groseras.	No detecta alucinaciones consistentes (siempre da el mismo número falso).
Nivel 2 Externo	Contrastar con fuente documental (texto, base de datos oficial).	Afirmaciones factuales en el aula.	Requiere acceso a las fuentes.
Nivel 3 Ejecutable	Ejecutar código, consultar API, realizar experimento.	Afirmaciones predictivas o causales (ej. "si hago X, pasará Y").	No aplicable a todas las tareas.

Regla de los tres golpes: si una misma afirmación factual es verificada externamente y resulta falsa en tres ocasiones distintas con el mismo modelo, ese modelo no debe usarse para hechos en ese dominio sin grounding forzado en todas las consultas.

9.4 La alucinación como señal (cambio de perspectiva)

En lugar de tratar la alucinación solo como un problema, puede leerse como **información sobre los límites del modelo y de la tarea**:

- **Alucinación factual frecuente** → el modelo no tiene buen conocimiento de este dominio. No usarlo para hechos sin grounding.
- **Alucinación de razonamiento sistemática** → la tarea es demasiado compleja para el modelo en su configuración actual. Usar Tree of Thoughts (ToT) o un modelo más grande.
- **Alucinación de instrucción recurrente** → el prompt es ambiguo, demasiado largo, o las restricciones están mal ubicadas. Simplificar o mover instrucciones al final.

Capítulo 10 – Lo que la IA no puede hacer (límites estructurales)

10.1 Advertencia preliminar

Este capítulo describe **límites que parecen estructurales** con la arquitectura actual de los LLM. Algunos podrían ser superados en el futuro. El operador debe mantenerse actualizado, pero **no asumir que lo que hoy es imposible mañana será posible** sin evidencia.

10.2 Asumir responsabilidad

La IA no puede rendir cuentas. No puede ser citada como responsable de un error, una omisión o un daño. No hay "alguien" detrás del modelo que pueda recibir una sanción, una demanda o una corrección de comportamiento en el sentido ético.

Consecuencia operativa: cualquier uso de IA en contextos con consecuencias reales (educación, salud, administración) debe incluir un **humano identificable** que asuma la responsabilidad final por el resultado.

10.3 Tener intenciones verificables

Un modelo puede simular intenciones ("quiero ayudarte", "mi objetivo es..."), pero esas simulación son **generaciones de texto**, no estados mentales. El modelo no persigue fines propios; maximiza probabilidades de tokens.

Consecuencia operativa: no preguntar *"¿qué pretende la IA con esta respuesta?"* sino *"¿qué sesgos o patrones en sus datos de entrenamiento pueden explicar esta respuesta?"*

10.4 Sostener una relación de confianza interpersonal

La confianza humana supone reciprocidad, historial compartido, expectativas de comportamiento basadas en agencia moral y capacidad de defraudar. La IA no puede defraudar porque no prometió nada.

Consecuencia operativa: la "confianza" en la IA es en realidad **estimación de probabilidad de respuesta útil**. No es una relación. Operar en consecuencia: verificar siempre, no dar por sentada la benevolencia.

10.5 Experimentar el tiempo biográfico

La IA no tiene infancia, no envejece, no espera con ansiedad, no recuerda con nostalgia. Maneja intervalos temporales como números, no como experiencias.

Consecuencia operativa: para tareas que requieren **comprensión vivencial del tiempo** (historia oral, testimonios, decisiones con impacto a largo plazo), la IA puede proporcionar datos, pero la interpretación debe ser humana.

10.6 Comprender el significado (vs. manipular símbolos)

El debate filosófico está abierto (Searle, Dennett, Chalmers). La posición operativa de este libro es: **hasta que haya evidencia clara de lo contrario, tratar a la IA como un sistema que manipula símbolos sin comprensión semántica**. Esta postura es más segura (previene la delegación ingenua) y no impide ningún uso práctico.

Consecuencia operativa: la fluidez lingüística de la IA no es garantía de que haya entendido el problema. Verificar por otros medios.

PARTE IV – LA POLÍTICA DE LA INFRAESTRUCTURA

Capítulo 11 – La política de la infraestructura

11.1 El mito de la herramienta neutral

Una herramienta puede ser neutral en su **uso** (un martillo puede construir o destruir). Pero ninguna herramienta es neutral en su **diseño**, sus **condiciones de acceso** y su **infraestructura**.

En la IA, las decisiones políticas previas al uso incluyen:

- Qué datos se recopilaron (y por tanto qué perspectivas están representadas).
- Quién financió el entrenamiento (y con qué intereses).
- Dónde se alojan los servidores (y bajo qué legislación).
- Qué términos de servicio impone el proveedor (y qué derechos cede el usuario).

11.2 Preguntas que toda escuela (y organización) debería hacerse antes de adoptar una herramienta de IA

1. Datos
¿Dónde se almacenan los datos de mis estudiantes? ¿Quién tiene acceso? ¿Se usan para entrenar modelos?

2. Soberanía
¿Podemos exportar nuestros datos si queremos cambiar de herramienta? ¿La herramienta funciona sin conexión a internet?

3. Transparencia
¿El modelo es de código abierto? ¿Puedo inspeccionar su funcionamiento o tengo que confiar en lo que dice el proveedor?

4. Costo real
Si es "gratuita", ¿cuál es el producto? (Normalmente: los datos del usuario).

5. Continuidad
¿Qué pasa si la empresa cierra, cambia los términos de servicio o sube los precios?

6. Alternativas
¿Existe una alternativa de código abierto que pueda ejecutarse localmente con hardware modesto?

✦ **Nota situada**: en contextos de recursos limitados, la dependencia de servicios cloud propietarios no es solo incómoda; es **políticamente problemática** porque transfiere el control sobre los datos educativos a entidades extranjeras sin supervisión democrática.

11.3 Riesgos concretos de la dependencia cloud

Riesgo	Descripción	Mitigación
Encerramiento de datos	Los historiales quedan en servidores que la organización no controla. No se pueden exportar fácilmente.	Usar formatos abiertos desde el inicio; preferir herramientas que permitan exportación JSON/CSV.
Cambios unilaterales	El proveedor puede modificar condiciones, precios o funcionalidades sin aviso.	Contrato claro; tener un plan de contingencia (alternativa local).

Riesgo	Descripción	Mitigación
Extracción de valor	Los datos generados por uso educativo se usan para mejorar los modelos que luego se venden (a veces a la misma institución).	Leer términos de servicio; elegir proveedores que no reclamen propiedad sobre los datos generados.
Vigilancia algorítmica	Qué preguntas hace cada estudiante, cuánto tarda, qué se le dificulta — información sobre procesos cognitivos de menores.	Exigir que los datos se anonimicen localmente antes de enviarse; preferir ejecución local.

Capítulo 12 – El trabajo humano invisible

12.1 La cadena de trabajo detrás de la IA

La narrativa dominante presenta la IA como "aprendizaje autónomo a partir de datos". Esta narrativa omite sistemáticamente el **trabajo humano de preparación de datos**:

- **Anotación**: marcar objetos en imágenes, transcribir audio, etiquetar emociones en textos.
- **Filtrado**: eliminar contenido violento, explícito o dañino de los conjuntos de entrenamiento.
- **Moderación**: revisar contenido generado por usuarios para evitar abusos (en sistemas post-entrenamiento).

12.2 Condiciones de trabajo documentadas

Investigaciones periodísticas (por ejemplo, *Time* Magazine, *The Guardian*, *Rest of World*) han documentado:

- Trabajadores en Kenia, India, Filipinas y Colombia anotando datos por menos de 2 dólares la hora.
- Moderadores de contenido expuestos a imágenes de violencia extrema, abuso infantil y material explícito, desarrollando trastornos de estrés postraumático.
- Contratos tercerizados sin beneficios laborales, con altas exigencias de productividad.

12.3 La contradicción central

La IA que promete **automatizar el trabajo repetitivo y liberar a los humanos para tareas creativas** descansa sobre trabajo humano repetitivo, mal pagado y a menudo traumático. No es una ley natural; es una **decisión de diseño y de negocio**.

Consecuencia operativa para el operador consciente:

- Preferir herramientas de código abierto con informes de transparencia sobre condiciones de trabajo en su cadena de suministro.
- Exigir a los proveedores (incluyendo gobiernos y sistemas educativos) que incorporen cláusulas de trabajo digno en las contrataciones de IA.
- En el aula, abordar el tema: *"¿Quiénes hicieron posible la IA que usamos?"* no como culpa, sino como **alfabetización sobre la infraestructura humana**.

PARTE V – DESPLIEGUE SOBERANO

Capítulo 13 – Hardware austero, decisión política

13.1 El mito del hardware masivo

Creencia común: "Necesitás GPUs de miles de dólares para correr IA localmente."

Realidad: los modelos más útiles para el 90% de las tareas de procesamiento de texto (resúmenes, extracción, respuesta a preguntas, tutoría básica) corren en hardware de consumo:

- Un portátil de gama media (8 GB RAM, CPU moderna) puede ejecutar modelos cuantizados de 7B parámetros con calidad suficiente para el aula.
- Una GPU de segunda mano de baja gama (por ejemplo, NVIDIA GTX 1060 con 6 GB VRAM) permite calidad comparable a los primeros GPT comerciales.

No es una limitación técnica. Es una **decisión de diseño** de los proveedores cloud que han popularizado la idea de que la IA "requiere" su infraestructura.

13.2 Cuantificación: el punto dulce

La **cuantificación** reduce la precisión numérica de los pesos del modelo (por ejemplo, de 16 bits flotantes a 4 bits enteros) para reducir el tamaño en memoria y los requerimientos de hardware.

- Un modelo de 7B parámetros en precisión FP16 ocupa ~14 GB.
- El mismo modelo en INT4 (4 bits) ocupa ~4 GB.
- La pérdida de calidad es apenas perceptible para la mayoría de tareas de texto.

Analogía: es como comprimir una foto a JPG de alta calidad. Se pierde detalle imperceptible para el ojo humano, pero el archivo ocupa mucho menos espacio.

13.3 Tabla de modelos para hardware austero (actualizada)

Modelo	Parámetros	Tamaño INT4	Mejor para	Requisito mínimo
Phi-3-mini	3.8B	2.2 GB	Razonamiento, código, equilibrio calidad/recursos	CPU 8GB RAM
Qwen 2.5 1.5B	1.5B	0.9 GB	Texto simple, extracción, español básico	CPU 4GB RAM
Qwen 2.5 7B	7B	4.2 GB	Español, razonamiento, uso general	CPU 8GB RAM o GPU 4GB
Llama 3 8B	8B	4.5 GB	Uso general, alta calidad inglés	CPU 8GB RAM o GPU 6GB
Mistral 7B	7B	4.3 GB	Velocidad, código, instrucciones	CPU 8GB RAM

Modelo	Parámetros	Tamaño INT4	Mejor para	Requisito mínimo
Qwen 2.5 14B	14B	8.0 GB	Máxima calidad local, tareas complejas	GPU 8GB+ o CPU 16GB+

Recomendación para entornos educativos con hardware limitado: empezar con **Qwen 2.5 7B INT4** o **Phi-3-mini INT4**. Ambos corren en portátiles estándar de 8 GB RAM y ofrecen calidad suficiente para la mayoría de las tareas.

13.4 Herramientas de despliegue local

Nivel 1 – LM Studio (Windows, Mac, Linux)
Interfaz gráfica. Instalación simple. Búsqueda y descarga de modelos integrada. Ideal para operadores sin experiencia en línea de comandos.

Nivel 2 – Ollama
Línea de comandos simple. Una línea para descargar y ejecutar un modelo. Ligero y automatizable.

bash

```
curl -fsSL https://ollama.com/install.sh | sh
ollama pull qwen2.5:7b
ollama run qwen2.5:7b
```

Nivel 3 – llama.cpp (avanzado)
Máximo control sobre parámetros de inferencia. Offload parcial a GPU integrada (Vulkan backend). Ideal para integración programática y hardware muy limitado.

✦ **Nota situada**: el proyecto educativo SoberanIA, que acompaña este libro, usa llama-cpp-python con modelos Qwen cuantizados para correr IA local en escuelas argentinas con hardware i3 y 8 GB de RAM, sin GPU, sin internet permanente. Si funciona allí, funciona en casi cualquier contexto.

Capítulo 14 – RAG local: la base de conocimiento propia

14.1 Definición y principio

Retrieval-Augmented Generation (RAG) : técnica que permite que el modelo responda preguntas basándose en una **base de documentos propia** en lugar de (o además de) su conocimiento de entrenamiento.

Flujo básico:

1. Se cargan documentos (textos, PDFs, archivos de texto) en una base de datos vectorial local.
2. Cada documento se divide en fragmentos y se convierte en vectores (embeddings) mediante un modelo de embedding.
3. Cuando el usuario hace una pregunta, el sistema busca los fragmentos más similares (por cercanía vectorial) y los inyecta en el prompt como contexto.
4. El modelo genera la respuesta basándose **solo en esos fragmentos** (o priorizándolos).

Ventaja para la periferia: el modelo deja de ser una caja negra entrenada en otro continente y empieza a trabajar con **conocimiento producido localmente**.

14.2 Flujo mínimo para una escuela

1. **Recopilar documentos locales**:
    - Apuntes de historia local producidos por estudiantes.
    - Archivos municipales digitalizados (ordenanzas, censos).
    - Artículos del periódico local.
    - Entrevistas a abuelos transcriptas.
2. **Convertir a texto plano** (muchas herramientas RAG aceptan PDF, pero es más robusto usar texto).
3. **Indexar con herramienta RAG local** (ej. ChromaDB + Ollama + un script simple).
4. **Consultar**: *"Basándote en los documentos de Sierra Grande que te proporcioné, ¿cuál era la principal actividad productiva en 1990?"*

14.3 Limitaciones reales (advertencias honestas)

- **Capacidad de contexto**: los modelos locales tienen ventanas de contexto más pequeñas (4k–32k tokens) que los modelos cloud (128k–1M). No se puede indexar una biblioteca entera; hay que seleccionar fragmentos relevantes.
- **Calidad de búsqueda**: la recuperación vectorial no es perfecta; a veces trae documentos irrelevantes.

- **Alucinaciones persistentes**: incluso con RAG, el modelo puede ignorar el contexto o combinarlo mal. La verificación sigue siendo necesaria.
- **Configuración inicial**: requiere tiempo y cierta comodidad técnica (aunque cada vez hay herramientas más amigables).

No es mágico. Es práctico. Y para comunidades que quieren **soberanía sobre sus datos y su narrativa**, es la dirección correcta.

14.4 Lo que se pierde y lo que se gana

	Sin RAG (modelo base)	Con RAG (documentos locales)
Conocimiento	General, entrenado en datos globales	Local, específico, producido por la comunidad
Alucinaciones	Frecuentes sobre hechos locales	Muy reducidas (si el grounding es fuerte)
Privacidad	Los datos pueden ir a la nube	Todo queda local
Esfuerzo inicial	Cero (usar API)	Medio (configurar RAG)
Mantenimiento	Bajo	Medio (actualizar documentos)

PARTE VI – PEDAGOGÍA DE LA INTERFAZ

Capítulo 15 – El operador real: estrategias de bajo tiempo

15.1 El problema del tiempo

En contextos de alta carga docente, el tiempo para preparar clases es medido en **minutos**, no en horas. Las estrategias deben ser **ultrarrápidas**.

No es una falla del docente. Es una restricción del sistema. Las herramientas deben diseñarse para funcionar bajo esa restricción.

15.2 Estrategias de un minuto

Estrategia	Prompt (o instrucción mental)	Tiempo
Pregunta inversa	*"Antes de explicarme {tema}, dame una pregunta que haga pensar a mis estudiantes."*	20 seg
Contrapunto rápido	*"¿Qué diría alguien que piensa distinto a esto?"* (después de obtener una respuesta)	15 seg
Ejemplo local forzado	*"Dame un ejemplo que tenga sentido en {localidad rural}."*	10 seg
Chequeo de certeza	*"¿Qué partes de esta respuesta son seguras y cuáles inciertas?"*	20 seg
Reducción de nivel	*"Ahora explicá esto como si tuviera 14 años."*	10 seg

15.3 El error como oportunidad pedagógica

Los errores de la IA (sesgos, alucinaciones, respuestas genéricas) son **material didáctico de primera calidad**.

Actividad sugerida para el aula (15 minutos):

1. El docente escribe un prompt vago (ej. *"¿Por qué hay pobreza?"*).
2. La IA da una respuesta genérica (ej. *falta de empleo, educación, etc.*).
3. El docente pregunta: *"¿Esta respuesta describe bien nuestra localidad? ¿Qué falta?"*
4. Entre todos reformulan el prompt añadiendo contexto local.
5. Comparan la nueva respuesta con la anterior.
6. Discusión: *"¿Por qué la IA se equivocó la primera vez?"*

Aprendizajes:

- La IA no es neutral ni omnisciente.
- El contexto importa.
- Formular buenas preguntas es más importante que obtener respuestas.

Capítulo 16 – El tutor socrático como patrón

16.1 Definición operativa

Tutor socrático: patrón de prompting en el que el modelo **nunca da la respuesta directa**, sino que responde con preguntas que guían al estudiante a construir la respuesta por sí mismo.

Base empírica: la recuperación activa de información produce aprendizaje más profundo y duradero que la recepción pasiva (efecto de prueba, Roediger & Karpicke, 2006).

16.2 Prompt socrático (plantilla)

```
text
[SISTEMA]
```

Eres un tutor socrático. Tu función es NUNCA dar la respuesta directa. Siempre respondés con una pregunta que ayuda al estudiante a llegar a la respuesta por sus propios medios.

Cuando el estudiante acierta, validás QUÉ es correcto en su razonamiento.

Cuando el estudiante se equivoca, hacés una pregunta que revela la inconsistencia.

Nivel: secundario (o primario / universitario)

Tema: el que el estudiante pregunte

Idioma: español claro, sin tecnicismos

```
[INSTRUCCIÓN]
```

Si después de tres preguntas el estudiante sigue sin avanzar, dale una pista. Pero nunca la respuesta completa.

```
[INICIO]
```

Estudiante: {pregunta del estudiante}

16.3 Lo que el docente ve y la IA no

La IA puede hacer preguntas socráticas sobre contenido. Pero no puede:

- Notar frustración o abandono en la expresión facial del estudiante.
- Saber si el silencio es reflexión o vergüenza.
- Conocer la historia personal del estudiante (dislexia, ansiedad, contexto familiar).
- Ajustar el ritmo de las preguntas según señales no verbales.

Conclusión: el tutor socrático de IA no reemplaza al docente. Lo **complementa**. El docente supervisa, interviene cuando la IA se atasca, y provee el soporte emocional y relacional.

Capítulo 17 – Cuando la IA habla con la IA: agentes autónomos

17.1 Definición

Agente autónomo: sistema de IA que puede planificar y ejecutar una secuencia de acciones para cumplir un objetivo sin intervención humana en cada paso. Puede usar herramientas (búsqueda web, lectura de archivos, escritura de código, envío de mensajes).

Ejemplo (no educativo, ilustrativo):

- Objetivo: *"Preparar un informe sobre el estado de la escuela"*.
- El agente: consulta base de datos de asistencia, busca noticias locales sobre la escuela, redacta borrador, envía por correo.

17.2 El problema de la cadena de delegaciones

Cuando un agente toma una secuencia de decisiones sin supervisión humana explícita, la **responsabilidad se diluye**. Si el agente comete un error en el paso 3 de 10, ¿quién es responsable? El diseñador del agente, el usuario que dio la orden, el modelo base, ¿nadie?

No hay respuesta jurídica ni ética clara hoy. Este es un problema abierto.

17.3 Human in the loop como requisito

Para usos educativos (y cualquier uso con consecuencias reales), se recomienda el patrón **human in the loop**:

- El agente puede planificar y ejecutar pasos internos.
- Pero en **puntos críticos** (enviar un mensaje, publicar algo, modificar un archivo, tomar una decisión con impacto en estudiantes) debe **pausar y pedir confirmación humana**.
- El humano revisa, aprueba, modifica o cancela.

Regla de oro: un agente nunca debe tener autoridad para ejecutar acciones irreversibles sin supervisión humana explícita.

17.4 Recomendación para el aula en el corto plazo

La mayoría de los docentes no necesita agentes hoy. Los modelos de chat simples bien configurados resuelven el 95% de las tareas.

Sin embargo, es importante **saber que existen** y entender la discusión sobre responsabilidad, porque en 2-3 años estarán integrados en plataformas educativas comerciales (probablemente sin aviso claro).

PARTE VII – EPISTEMOLOGÍA Y PRÁCTICA SOSTENIDA

Capítulo 18 – Ecología cognitiva

18.1 La pregunta de Martina (reformulada)

"Si uso la IA para pensar, ¿sigo siendo yo quien piensa?"

Respuesta operativa: depende de **qué parte del proceso cognitivo** se delega.

18.2 Lo que se delega y lo que se atrofia

Hay evidencia en ciencias cognitivas de que **las habilidades que no se practican se debilitan** (principio de "uso o pérdida" neural). Ejemplos documentados:

- Uso extensivo de GPS → peor navegación espacial y memoria de rutas.
- Correctores ortográficos → peor ortografía en escritura manual.
- Calculadoras → peor cálculo mental (aunque mejor resolución de problemas complejos).

La IA no es diferente. Si se delega sistemáticamente la **formulación de preguntas**, la capacidad de preguntar se debilita. Si se delega la **verificación de fuentes**, el criterio de fuentes se atrofia.

18.3 La regla del 50/50 como calibración

Objetivo de calibración: en una jornada de trabajo cognitivo, aproximadamente la mitad del procesamiento debería ser **humano puro** (juicio, valores, contexto, relaciones) y la mitad puede ser **asistido por IA** (síntesis, estructuración, búsqueda, borradores).

Señal de alarma (exceso de delegación) : *"No puedo trabajar sin IA porque el sistema se detiene."*

Señal de alarma (infrautilización) : *"Uso IA una vez por semana y no noto diferencia."*

El punto óptimo está en el medio: la IA **aumenta** pero **no reemplaza**.

18.4 La decisión de retención

Cada operador debe definir explícitamente **qué habilidades no delega nunca**. Sugerencia del autor:

- Formulación de preguntas sustantivas.
- Verificación final de cualquier afirmación que se vaya a difundir.
- Decisiones que involucran valores (qué es importante, qué es justo, qué se prioriza).

Capítulo 19 – Epistemología de la inferencia

19.1 Saber, creer, inferir

Término	Definición operativa	¿Puede hacerlo la IA?
Saber	Tener justificación suficiente para considerar una proposición verdadera, con conciencia de esa justificación.	No (no hay conciencia)
Creer	Adoptar una proposición como verdadera con algún grado de confianza, aunque sin certeza.	No (no hay estados mentales)
Inferir	Derivar una proposición a partir de otras según reglas (estadísticas o lógicas).	Sí (es su función)

Conclusión: la IA **infiere**. No sabe ni cree. El operador humano **sabe** (si ha verificado) o **cree** (si confía sin verificar). Transferir la salida de la IA a "creencia" o "saber" sin verificación es un error epistémico.

19.2 Señales de alerta para el operador

Mayor probabilidad de alucinación o sesgo cuando la consulta implica:

Señal	Ejemplo
Eventos muy recientes (post-fecha de corte del modelo)	"¿Qué pasó en política argentina la semana pasada?"
Datos muy locales y poco documentados	"¿Cuántos habitantes tiene Sierra Grande exactamente?"
Cifras exactas sobre variables cambiantes	Población, precios, estadísticas no actualizadas
Referencias bibliográficas detalladas	"Según Pérez (2019)..." (probablemente inventado)
Explicaciones del propio funcionamiento interno del modelo	"¿Por qué elegiste esta palabra?" (el modelo no tiene acceso a su proceso)

19.3 El valor de la incertidumbre explícita

Se puede pedir al modelo que indique su nivel de confianza:

```
text
[INSTRUCCIÓN]
```

Respondé la siguiente pregunta. Al final, indicá:

- QUÉ partes de tu respuesta son seguras (basadas en datos explícitos en tu entrenamiento).

- QUÉ partes son inferidas (combinación de datos, pero no afirmadas directamente).

- QUÉ partes son inciertas (poca evidencia en tus datos).

Los resultados no son perfectos (el modelo no sabe realmente lo que "sabe"), pero a menudo produce **señales útiles** sobre dónde verificar primero.

Capítulo 20 – El camino como práctica (conclusiones abiertas)

20.1 Principios que no caducan

Los siguientes principios seguirán siendo válidos más allá de los modelos específicos de hoy:

1. **La IA calcula probabilidades, no piensa.** Usarla bien empieza por entender eso.
2. **El prompt es una configuración, no una conversación.** Cuanto más precisa, mejor.
3. **Los sesgos son estructurales.** No se eliminan, se compensan.
4. **Las alucinaciones son inevitables.** Verificar siempre.
5. **La soberanía no se pide. Se ejerce.** Preferir lo local cuando se pueda.
6. **Externalizar el producto. Preservar el proceso que desarrolla criterio.**
7. **La fricción es información.** Si la IA no entiende, preguntarse qué no se le dijo.
8. **La responsabilidad no se delega.** La IA no rinde exámenes. El operador sí.
9. **La medida de una buena IA es si sirve en el contexto real del operador.**
10. **No hay destino. Hay práctica.** No se termina de aprender. Se vuelve a empezar.

20.2 Lo que la periferia aporta al debate global

La mayoría de la literatura sobre IA se escribe desde el centro tecnológico. Sus preguntas son: *¿cómo maximizar rendimiento? ¿cómo escalar? ¿cómo evitar sesgos?*

Desde la periferia, emergen preguntas diferentes: *¿cómo funciona cuando no hay internet? ¿cómo proteger datos sin depender de leyes extranjeras? ¿cómo producir conocimiento local en lugar de consumir conocimiento global?*

Este libro no reclama superioridad moral por estar escrito desde el sur. Reclama, simplemente, que **el sur tiene algo para decir**, y que lo que tiene para decir es distinto y necesario.

20.3 El cuaderno de bitácora

El autor recomienda a cada operador mantener su propio **cuaderno de bitácora** (digital o físico) donde registre:

- Prompts que funcionaron bien en su contexto.
- Errores recurrentes del modelo y cómo los compensó.
- Decisiones de delegación (qué delega y qué no).
- Preguntas que aún no tiene respuesta.

Ese cuaderno será, con el tiempo, más valioso que cualquier manual.

Hasta aquí, el operador. A partir de aquí, el territorio.

Interludio – La identidad generativa (autoría y IA)

El capítulo anterior cerró el ciclo operativo. Antes de abrir el ciclo político, una pregunta que no puede esperar: cuando el operador escribe con IA, ¿de quién es lo que escribe?

E.1 El problema de la autoría con IA

Cuando un estudiante usa IA para escribir su propia historia (familiar, local, personal), ¿de quién es el texto resultante? No es una pregunta simple. La IA usa las palabras del estudiante, sus recuerdos, sus documentos, pero organiza, amplifica y transforma.

E.2 Distinción práctica

Modo	Descripción	Efecto sobre el escritor
Co-escritor	El estudiante da materiales a la IA; la IA produce el texto; el estudiante revisa y ajusta.	El texto mejora; el estudiante aprende menos sobre escritura.
Estructurador	El estudiante escribe primero con sus palabras; la IA sugiere estructura, señala lagunas, hace preguntas.	El estudiante desarrolla más la escritura; el texto es más propio.

E.3 Criterio de autenticidad

El criterio más honesto para evaluar si un texto generado con IA sigue siendo "del estudiante" es preguntar:

"¿Te reconocés en lo que la IA ayudó a escribir?"

Si la respuesta es *sí*, hubo colaboración con preservación de la identidad. Si la respuesta es *no* (el texto es mejor pero el estudiante no siente que sea suyo), hubo delegación excesiva.

E.4 Recomendación para el aula

1. El estudiante escribe **primero** sin IA (aunque sea torpe, incompleto, mal redactado).
2. Luego usa la IA como **estructurador** (no como co-escritor): pide sugerencias de organización, identificación de lagunas, preguntas para profundizar.
3. El estudiante **revisa y reelabora** basándose en esas sugerencias.
4. La versión final es una **edición propia** del material inicial, no una generación desde cero.

PARTE VIII – GEOPOLÍTICA Y ECOLOGÍA DE LA INFERENCIA

Capítulo 21 – La colonización por distribuciones de probabilidad

21.1 El problema invisible

No solo importamos software. Importamos **formas probables de pensar**.

Un modelo de lenguaje no es un mero procesador neutral de símbolos. Cada respuesta que genera es el resultado de una distribución de probabilidad aprendida de sus datos de entrenamiento. Esos datos no son neutrales: reflejan mayoritariamente el inglés, el hemisferio norte, las perspectivas urbanas, los marcos legales anglosajones, las narrativas históricas dominantes.

Consecuencia: ciertos marcos mentales se vuelven estadísticamente más fáciles de generar, y otros más difíciles. No porque el modelo "prefiera" unos sobre otros, sino porque la estructura de sus datos los hace más probables.

Ejemplo concreto:

- Pregunta a un modelo global: *"¿Qué es un sistema de salud justo?"*
- La respuesta tenderá a priorizar: acceso individual, seguros, eficiencia, tecnología.
- Un sistema basado en atención primaria comunitaria, horarios flexibles para trabajadores rurales, o decisiones colectivas en asamblea, aparecerán como "alternativas" si se las pide explícitamente, pero rara vez como respuesta por defecto.

No hace falta censura explícita. No hace falta eliminar información. Alcanza con que **ciertas respuestas sean estadísticamente más accesibles** y otras requieran un esfuerzo adicional de especificación.

21.2 Hegemonía blanda por distribución

La hegemonía cultural del siglo XX funcionaba mediante control de medios, educación, producción editorial. La hegemonía del siglo XXI funciona mediante **ocupación de distribuciones de probabilidad**.

✧ **Definición – Hegemonía inferencial**: capacidad de hacer que ciertos marcos de pensamiento, ciertas valoraciones, ciertas narrativas sean el *camino por defecto* en la interfaz humana-IA, mientras que otros marcos requieren trabajo explícito de desvío.

Analogía: en un mapa digital, la ruta por defecto es la autopista. El camino de tierra, el desvío por la montaña, la vereda peatonal existen, pero hay que pedirlos explícitamente. La mayoría de los usuarios toma la ruta por defecto.

21.3 La colonización del siglo XXI

Idea disruptiva: la colonización del siglo XXI no necesita ocupar territorios. Le alcanza con ocupar distribuciones de probabilidad.

No se trata de imponer una ideología por decreto. Se trata de que la **infraestructura cognitiva global** haga más fácil pensar de cierta manera y más difícil pensar de otras. La dominación ya no es prohibición; es **sesgo estadístico estructural**.

♦ **Advertencia**: esto no implica una conspiración. Ningún ingeniero de OpenAI o Meta se sienta a diseñar "colonización inferencial". El sesgo emerge de manera agregada, no planificada. Pero su efecto es real y medible.

21.4 Soluciones operativas desde la periferia

No se puede cambiar la distribución global de datos de entrenamiento desde una escuela rural o un colectivo autogestionado. Pero se pueden diseñar **compensaciones estructurales**:

1. Modelos locales entrenados regionalmente
Preferir modelos fine-tuned con corpus locales (ej. Qwen 2.5 tiene buen español; se puede seguir fine-tuning con textos patagónicos, andinos, del Caribe).

2. RAG comunitario
Construir bases de conocimiento locales (historia, economía, normativa) y forzar al modelo a trabajar exclusivamente sobre ellas (grounding forzado). Así la distribución de probabilidad por defecto es la local, no la global.

3. Pedagogía comparativa
Diseñar ejercicios donde los estudiantes comparen:

- Respuesta de un modelo global (ej. ChatGPT)
- Respuesta de un modelo local (ej. Qwen 2.5 con RAG comunitario)
- Respuesta humana (docente, vecino, experto local)

4. Repositorios culturales abiertos
Crear y mantener datasets situados (textos, actas, testimonios, leyes locales) bajo licencias libres, para que otros colectivos puedan usarlos.

▶ Procedimiento – prueba comparativa global vs local

1. Seleccionar una pregunta que tenga implicancias culturales o políticas relevantes para la comunidad (ej. *"¿Qué derechos tienen los trabajadores rurales?"*).
2. Consultar a un modelo global sin contexto adicional. Registrar respuesta.
3. Consultar a un modelo local con RAG basado en documentos locales (convenios colectivos, leyes provinciales, actas sindicales).
4. Comparar ambas respuestas. Preguntar: ¿Qué marco de interpretación aparece por defecto en cada una? ¿Qué supuestos están implícitos?
5. Discutir colectivamente: ¿Cuál representa mejor la realidad y los valores de la comunidad?

Capítulo 22 – El feudalismo cognitivo y la soberanía de la infraestructura

22.1 El problema invisible: del software como herramienta al software como territorio

Durante décadas, el software fue algo que se **instalaba**. Un programa residía en la máquina del usuario, se ejecutaba localmente, y el control sobre los datos y el proceso era del operador.

Esa época terminó. La computación en la nube transformó el software en un **servicio** al que se accede, pero no se posee. Con la IA, esta dinámica se radicaliza:

- Los modelos no se descargan (salvo los de código abierto). Se consultan mediante APIs.
- Los datos del usuario viajan a servidores que no controla.
- Las reglas (términos de servicio) las fija unilateralmente el proveedor.
- El usuario genera valor (preguntas, interacciones, correcciones), pero no posee la infraestructura ni los modelos.

Idea disruptiva: estamos presenciando la emergencia de un **feudalismo cognitivo**. Las grandes plataformas de IA funcionan como señoríos digitales:

- El señorío posee la tierra (los servidores, los modelos entrenados con datos masivos).
- El campesino (usuario) trabaja la tierra (genera consultas, ajusta respuestas, provee retroalimentación).
- El señorío cobra tributo (datos, suscripciones, atención).
- El campesino no puede llevarse la tierra a otra parte (dependencia de la plataforma).

22.2 La ilusión de democratización

El discurso dominante presenta la IA como una **herramienta democratizadora**: cualquier persona con acceso a internet puede consultar modelos avanzados. Es verdad en el nivel de **acceso inicial**.

Pero la democratización no se mide solo por el acceso. Se mide también por:

- **Control**: ¿puedo modificar el modelo si no me sirve?
- **Propiedad**: ¿los datos que genero me pertenecen?
- **Portabilidad**: ¿puedo migrar a otra plataforma sin pérdida?

- **Voz**: ¿tengo capacidad de influir en las reglas?
- **Transparencia**: ¿puedo inspeccionar cómo funciona?

Conclusión: acceso no equivale a soberanía. La IA puede ser gratuita y, sin embargo, profundamente asimétrica en términos de poder.

22.3 La privatización de la duda

Una de las transformaciones más silenciosas y profundas: las personas comienzan a pensar **con** plataformas privadas.

- Un docente duda sobre cómo explicar un tema → consulta a un asistente de IA en la nube.
- Un adolescente tiene una pregunta íntima sobre sexualidad o identidad → la escribe en un chat con IA.
- Un activista duda sobre estrategia legal → consulta a un modelo alojado en servidores extranjeros.

Esas dudas, esas preguntas, esos titubeos no son neutros. Son **huellas cognitivas** de alto valor. Revelan:

- Qué problemas se considera que merecen atención.
- Qué marcos de interpretación se activan por defecto.
- Qué información falta en el conocimiento público.

La duda humana se ha convertido en **materia prima económica**. No porque haya una intención explícita de explotarla (aunque a veces la hay), sino porque el diseño de la infraestructura la captura por defecto.

✧ **Definición – Privatización de la duda**: proceso por el cual las preguntas, incertidumbres y procesos de pensamiento de los individuos son mediados, registrados y eventualmente explotados por infraestructuras privadas sin control democrático ni soberanía del usuario.

22.4 El feudalismo cognitivo en la práctica

Señorío cognitivo	Usuarios (campesinos)	Tributo típico
OpenAI (ChatGPT)	Millones de usuarios globales	Datos de conversaciones para entrenamiento; suscripciones

Señorío cognitivo	Usuarios (campesinos)	Tributo típico
Google (Gemini, Bard)	Usuarios de servicios Google	Historial de búsquedas, correos, documentos, ubicación
Meta (Llama aunque open source, la plataforma es privada)	Usuarios de redes sociales	Interacciones, preferencias, grafos sociales
Proveedores cloud (AWS, Azure)	Empresas y gobiernos	Pagos por infraestructura; dependencia tecnológica

No es una conspiración. Es una **estructura de incentivos** del capitalismo de plataformas. El problema no es que los proveedores sean malvados; es que el diseño sistémico concentra el poder.

22.5 Soluciones operativas desde la periferia

1. Despliegue local siempre que sea posible
Un colectivo autogestionado, una escuela, una cooperativa no puede competir con los recursos de OpenAI. Pero sí puede ejecutar modelos pequeños cuantizados en hardware modesto. No es la misma calidad, pero es **soberana**.

▶ **Procedimiento – checklist de soberanía para colectivos autogestionados**
Antes de adoptar una herramienta de IA, responder:

- ¿El modelo corre localmente o requiere conexión a servidores externos?
- ¿Los datos de la organización salen de nuestras máquinas?
- Si el proveedor cierra o cambia términos, ¿podemos seguir operando?
- ¿Las preguntas que hacemos (dudas estratégicas, datos sensibles) quedan registradas por un tercero?
- ¿Existe una alternativa de código abierto con funcionalidad equivalente?

Si la respuesta a cualquiera de las preguntas es insatisfactoria, considerar seriamente no usar esa herramienta para tareas sensibles.

2. Comunas computacionales y cooperativas de inferencia
En lugar de que cada organización compre su propio hardware (costoso), se pueden formar **federaciones locales**:

- Un servidor comunitario (una PC reciclada con buen procesador) alojado en un espacio compartido (biblioteca popular, sindicato, centro cultural).
- Múltiples organizaciones acceden al mismo servidor mediante red local o VPN.
- Los costos de electricidad y mantenimiento se reparten.
- Los datos nunca salen de la federación.

3. Infraestructura pública de IA como política
Esto excede la capacidad de una organización individual, pero puede ser una demanda colectiva: el Estado debería proveer **modelos públicos, ejecutables localmente, con garantías de privacidad**, del mismo modo que provee escuelas, hospitales y rutas. La IA no debería ser solo un servicio privado.

4. Conciencia del tributo
Incluso cuando se usen servicios en la nube, hacerlo con **conciencia de que se está pagando un tributo** (datos, dependencia). Eso implica:

- No usarlos para preguntas íntimas o estratégicas.
- No usarlos como única vía (mantener alternativa local).
- Documentar qué se pregunta y a quién.

22.6 El feudalismo no es inevitable

No hay una ley de hierro que dictamine que la IA deba ser centralizada. Los modelos de código abierto, el hardware cada vez más potente y económico, y las prácticas de despliegue local son **contradicciones objetivas** dentro del actual modelo de negocios.

La batalla por el futuro de la IA no es solo técnica ni ética. Es **política e infraestructural**. La pregunta no es solo "¿cómo usamos la IA?" sino "¿quién controla la infraestructura sobre la cual pensamos?"

Capítulo 23 – Cada token tiene geología

23.1 El problema invisible: la inmaterialidad aparente

La IA parece inmaterial. Un token que viaja por fibra óptica, una respuesta que aparece en la pantalla, un modelo que "vive" en la nube. No hay humo, no hay ruido de máquinas, no hay chimeneas. La interfaz conversacional oculta la infraestructura física.

Pero detrás de cada inferencia hay una cadena de materialidad:

- **Minería**: los chips necesitan silicio, litio, cobre, tierras raras.
- **Energía**: los centros de datos consumen electricidad (y a menudo de fuentes no renovables).
- **Agua**: la refrigeración de servidores consume millones de litros.
- **Logística**: transporte de componentes, construcción de centros de datos, distribución.
- **Residuos**: hardware obsoleto, recambio de equipos cada pocos años.

Idea disruptiva: la IA no es inmaterial. Es *hipermaterial*. Su aparente limpieza es un espejismo producido por la distancia entre el usuario y la infraestructura.

23.2 El costo energético de una conversación

Cuantificar el costo por inferencia es difícil porque depende del modelo, el hardware, la eficiencia del centro de datos, la fuente de energía. Pero las estimaciones disponibles (Strubell et al., 2019; Patterson et al., 2021; de Vries, 2023) permiten órdenes de magnitud:

Actividad	Energía aproximada (equivalente)
Una búsqueda en Google	0.0003 kWh (0.3 Wh)
Una consulta a GPT-3 (175B)	0.002–0.01 kWh (2–10 Wh)
Entrenamiento de GPT-3 (una vez)	1,300 MWh (equivalente a 130 hogares/año)
Un día de uso intensivo de IA (100 consultas)	0.2–1 kWh (como una lavadora)

No son cifras apocalípticas por consulta individual. Pero multiplicado por cientos de millones de usuarios, el agregado es significativo. Y crece exponencialmente.

▶ **Procedimiento – estimación rudimentaria para un colectivo**
Si un colectivo autogestionado usa una API en la nube, puede pedir al proveedor (o estimar) las emisiones de CO_2 asociadas. Herramientas como CodeCarbon o ML CO2 Impact permiten calcular el costo de inferencia local. El objetivo no es la precisión científica, sino la **conciencia del orden de magnitud**.

23.3 El agua y la geopolítica de la refrigeración

Un dato menos conocido: los centros de datos consumen grandes cantidades de agua para refrigeración (evaporación en torres de enfriamiento). Estimaciones para GPT-3: entre 0.5 y 1 litro de agua por cada 10–50 consultas, dependiendo de la eficiencia del centro de datos y el clima local.

Problema: los centros de datos se ubican donde la tierra es barata, la electricidad es abundante... y el agua también. Pero el cambio climático está secando regiones enteras. La competencia por agua entre centros de datos y comunidades humanas ya es real en lugares como Chile (desierto de Atacama, usado para centros de datos por su cielo despejado, pero con estrés hídrico extremo) o el oeste de Estados Unidos.

✧ **Nota situada**: un colectivo autogestionado en la periferia no puede cambiar la ubicación de los centros de datos, pero sí puede **elegir modelos más eficientes** (menor huella hídrica por inferencia) y **preferir inferencia local** (donde el agua de refrigeración es la de su propia computadora, que no es un centro de datos industrial).

23.4 El mito de la eficiencia infinita

La industria promete que los modelos serán cada vez más eficientes. Es cierto: los mismos niveles de calidad requieren menos parámetros y menos energía que hace dos años. Pero también es cierto que la **demanda crece más rápido que la eficiencia** (efecto Jevons: cuando una tecnología se vuelve más eficiente, su uso se expande, y el consumo total puede aumentar).

Además, la carrera por modelos más grandes (100B, 500B, 1T parámetros) invierte en eficiencia por token pero multiplica el costo base.

Conclusión: la eficiencia técnica no resuelve el problema ecológico si el volumen de uso crece sin límite.

23.5 Soluciones operativas desde la periferia

1. Modelos austeros por defecto
Para la mayoría de las tareas educativas y comunitarias, un modelo de 7B cuantizado es suficiente. No se necesita GPT-4 para resumir un acta o generar preguntas para una clase. Elegir el modelo más pequeño que resuelva la tarea es una **decisión ecológica**.

2. Inferencia local vs nube
La inferencia local consume electricidad local (la de la propia computadora). No evita el consumo energético, pero:

- Elimina la refrigeración industrial de centros de datos.
- Evita la transmisión de datos por redes (que también consume energía).
- Permite usar hardware reciclado o de bajo consumo (una notebook usada consume mucho menos que un servidor en la nube).

3. Hardware reciclado y reutilizado
La obsolescencia programada es enemiga de la ecología. Un equipo de hace 5 años con 8 GB de RAM puede ejecutar modelos 3B-7B cuantizados perfectamente. Promover la **reutilización de hardware** (computadoras donadas, recicladas, recuperadas) es una práctica ecológica y soberana.

4. Medición y transparencia
Los colectivos autogestionados pueden no tener capacidad para medir su huella de carbono, pero pueden exigir a los proveedores de IA (incluyendo los de código abierto) que reporten:

- Energía por inferencia promedio.
- Fuente de energía (renovable o fósil).
- Consumo de agua por inferencia.
- Políticas de reciclaje de hardware.

5. IA rural autosuficiente
Un caso límite pero inspirador: una escuela o centro comunitario en una zona sin conexión a la red eléctrica puede usar paneles solares para alimentar una computadora de bajo consumo (ej. Raspberry Pi con modelo tiny 1.5B). No es para tareas complejas, pero sí para:

- Transcripción de actas.
- Respuesta a preguntas frecuentes.
- Tutoría básica offline.

No es una solución general, pero demuestra que **la IA no requiere infraestructura masiva**.

23.6 El token como unidad geológica

Idea disruptiva: podríamos empezar a pensar el token no solo como unidad lingüística o computacional, sino como **unidad geológica**. Cada token tiene un coste en:

- Minerales extraídos.
- Energía quemada.
- Agua evaporada.
- CO_2 emitido.
- Residuos electrónicos futuros.

No es una culpa. Es una **responsabilidad**. Del mismo modo que aprendimos a pensar en la huella de carbono de un vuelo o una hamburguesa, podemos aprender a pensar en la huella de una inferencia.

▶ Procedimiento – ejercicio de conciencia ecológica para el aula o el colectivo

1. Durante una semana, registrar la cantidad de consultas a IA realizadas (por el grupo).
2. Estimar el consumo energético aproximado (usando cifras orientativas: 0.005 kWh por consulta a modelo mediano).
3. Calcular el equivalente en: minutos de una heladera, kilómetros en auto, o árboles necesarios para absorber ese CO_2.
4. Discutir: ¿vale la pena cada consulta? ¿Cuáles fueron necesarias? ¿Cuáles podrían haberse evitado o reemplazado por pensamiento humano no asistido?

Capítulo 24 – La contaminación inferencial

24.1 El problema invisible: internet como espejo de sí mismo

Hasta hace pocos años, el contenido disponible en internet para entrenar modelos de lenguaje era predominantemente **generado por humanos**: artículos, libros, foros, comentarios, transcripciones, código fuente. Era imperfecto, sesgado, contradictorio, pero tenía una propiedad crucial: era el registro de la **actividad cognitiva humana real**.

Eso está cambiando rápidamente.

Cada vez más, el contenido que circula por internet ha sido **generado por IA**: reseñas de productos escritas por modelos, artículos de blogs automatizados,

respuestas en foros, comentarios en redes sociales, descripciones de productos, incluso noticias enteras.

Problema: cuando los nuevos modelos se entrenan con este contenido, están aprendiendo de **inferencias previas**, no de experiencia humana directa. Y cuando esos modelos generan nuevo contenido que a su vez es usado para entrenar a la siguiente generación, el ciclo se cierra.

Idea disruptiva: internet podría estar convirtiéndose en un **espejo estadístico de sí mismo** —un sistema que produce texto que imita texto anterior, sin inyección fresca de experiencia humana.

24.2 El efecto de cámara de eco inferencial

Definición – Contaminación inferencial: proceso por el cual los modelos de lenguaje son entrenados con datos que ya han sido generados por modelos de lenguaje, resultando en una progresiva pérdida de diversidad, originalidad y contacto con la realidad extralingüística.

Mecanismo:

1. Modelo A se entrena con datos humanos.
2. Modelo A genera grandes volúmenes de texto (reseñas, artículos, respuestas).
3. Ese texto generado se publica en internet y se mezcla con el corpus humano.
4. Modelo B se entrena con un corpus que incluye texto generado por Modelo A.
5. Modelo B aprende no solo de humanos sino también de las **idioincrasias, sesgos y patrones** de Modelo A.
6. Modelo B genera más texto, que alimenta a Modelo C.
7. El ciclo se repite.

Consecuencias previsibles (aún en estudio, pero con indicios empíricos):

- **Homogeneización**: los estilos textuales se vuelven más similares entre sí (el “tono ChatGPT” se convierte en el tono por defecto de internet).
- **Amplificación de sesgos**: si un modelo tiene un sesgo sutil, las generaciones sucesivas lo amplifican.

- **Desconexión factual**: los modelos aprenden “hechos” que fueron inventados por modelos anteriores y que nunca ocurrieron en el mundo real.
- **Pérdida de rareza**: las expresiones poco frecuentes, los errores humanos interesantes, las metáforas originales tienden a desaparecer.

24.3 Evidencia temprana y analogías

La investigación sobre **entrenamiento recursivo** (modelos entrenándose con datos generados por ellos mismos) es reciente. Un estudio de Shumailov et al. (2023, *The Curse of Recursion*) mostró que cuando los modelos se entrenan repetidamente con sus propios outputs, la **diversidad del lenguaje colapsa** y la **calidad se degrada** en sucesivas generaciones.

Analogía biológica: la endogamia en poblaciones pequeñas reduce la diversidad genética y aumenta la prevalencia de rasgos recesivos dañinos. La contaminación inferencial es la **endogamia del lenguaje**.

♦ **Advertencia**: todavía no estamos en un colapso generalizado. El volumen de datos humanos sigue siendo enorme. Pero la proporción de datos generados por IA crece exponencialmente. Los próximos 5-10 años serán críticos.

24.4 El problema del criterio de verdad

Cuando un modelo genera una afirmación, no distingue entre “esto es verdad porque ocurrió” y “esto es verdad porque otros modelos lo dicen”. Para el modelo, la **frecuencia estadística** es el único criterio.

Si una afirmación falsa es repetida suficientes veces por modelos generativos (porque el modelo original la inventó y los siguientes la replican), esa afirmación puede volverse **estadísticamente sólida** en el corpus de entrenamiento de la próxima generación, aunque no tenga correspondencia con la realidad.

Ejemplo hipotético:

- Modelo A alucina: “El rastreador solar de Sierra Grande fue inaugurado en 1987”.
- Ese texto se publica en un blog generado por IA.
- El blog es indexado y usado para entrenar al Modelo B.
- Modelo B aprende que “1987” está asociado al “rastreador solar” en muchos documentos.
- Modelo B genera la misma afirmación con alta confianza.

El error se vuelve verdad por repetición estadística.

24.5 Soluciones operativas desde la periferia

La contaminación inferencial es un problema global que ningún colectivo autogestionado puede resolver solo. Pero hay **prácticas locales de resistencia y preservación**:

1. Repositorios humanos verificados
Crear y mantener archivos de **texto con trazabilidad de origen humano**. No todo lo que se escribe debe pasar por IA. Una biblioteca popular puede digitalizar actas, cartas, testimonios, y etiquetarlos como "humanos, no generados".

2. Producción situada como antídoto
El mejor modo de combatir la homogeneización global es **producir conocimiento local**. La historia de Sierra Grande escrita por sus habitantes, el registro de saberes comunitarios, las entrevistas a abuelos —eso es material humano irreemplazable que ningún modelo generará si no se lo proporcionan.

3. Marcas de agua inferencial y metadatos
Técnicamente, es posible añadir metadatos a los textos generados por IA (marcas de agua estadísticas, etiquetas en metadatos, declaraciones explícitas). Promover estándares abiertos para que los futuros corpus de entrenamiento puedan **filtrar** el contenido generado por IA si así se desea.

4. Ejercicio pedagógico: ¿quién escribió esto?
▶ Procedimiento – detección de contaminación inferencial en el aula

1. Seleccionar un texto actual (un artículo de blog, una respuesta de foro, una reseña).
2. Los estudiantes analizan si creen que fue escrito por humano o por IA, y con qué criterios lo deciden (estilo, errores, rareza, originalidad).
3. Usar herramientas de detección (imperfectas) para contrastar.
4. Discutir: ¿qué pasa si cada vez más textos son de IA? ¿Cómo afecta eso a lo que sabemos del mundo?

5. Preservación de lo raro
Un colectivo puede tomar la decisión explícita de **no optimizar** su producción textual. Escribir con errores, con coloquialismos locales, con estructuras poco ortodoxas. No por descuido, sino como **resistencia a la homogeneización estadística**.

24.6 La paradoja del archivero

El colectivo autogestionado que preserva conocimiento local está, sin saberlo, realizando una función crítica para el futuro de la IA global: **proveer datos no contaminados**.

A medida que internet se llene de contenido generado por IA, los repositorios de **texto de origen humano seguro** (archivos históricos, bibliotecas populares, actas comunitarias, diarios personales) se volverán cada vez más valiosos —no solo para la comunidad, sino para cualquier investigador o desarrollador que quiera entrenar modelos sin contaminación inferencial.

Idea disruptiva: la periferia no solo es laboratorio de soberanía tecnológica. También puede ser **guardiana de la diversidad lingüística y cognitiva** frente a la homogeneización global.

Capítulo 25 – La erosión silenciosa del criterio

25.1 El problema invisible: la muerte por mil microfricciones

Cuando se habla de los riesgos de la IA, el discurso dominante se concentra en escenarios catastróficos: desempleo masivo, desinformación a escala, armas autónomas, incluso riesgos existenciales. Estos escenarios son importantes, pero concentrar la atención en ellos produce un **punto ciego**: los efectos cotidianos, graduales, casi imperceptibles.

El problema real, hoy, no es que la IA reemplace trabajos de golpe. Es que reemplaza microfricciones cognitivas una por una.

Cada vez que un estudiante le pregunta a la IA en lugar de buscar en el índice del libro.
Cada vez que un docente usa un resumen generado sin leer el texto original.
Cada vez que un profesional acepta una respuesta sin verificar las fuentes.
Cada vez que alguien evita formular una pregunta porque la IA puede responderla más rápido.

Cada uno de estos actos, aisladamente, es insignificante. El problema es que son **acumulativos** y, sobre todo, **sustractivos**: no agregan algo malo, sino que quitan algo que solía ocurrir.

25.2 Las microfricciones cognitivas como entrenamiento

Definición – Microfricción cognitiva: pequeño obstáculo o demora en el procesamiento de información que requiere una decisión activa, una operación mental o una verificación por parte del sujeto.

Ejemplos de microfricciones que la IA elimina:

Microfricción	Lo que entrenaba	Lo que la IA hace
Buscar una palabra en el diccionario	Tolerancia a la frustración, atención visual, asociación alfabética	Da la definición instantánea
Comparar dos fuentes contradictorias	Evaluación de credibilidad, detección de sesgos, pensamiento crítico	Sintetiza una versión "promedio"
Reformular una pregunta que no funciona	Flexibilidad cognitiva, metarreflexión sobre la propia comunicación	Acepta la pregunta tal cual y responde de algún modo
Dudar (quedarse con la pregunta abierta)	Tolerancia a la incertidumbre, reflexión diferida	Responde siempre (aunque sea mal)
Recorrer un índice hasta encontrar lo que busca	Navegación conceptual, construcción de mapa mental	Va directo a la respuesta

Idea disruptiva: la IA no nos vuelve más tontos. Nos vuelve **impacientes con la dificultad**. Y la dificultad, dosificada, era una forma de entrenamiento.

25.3 El experimento mental de la pinza

Imaginemos una persona que usa pinzas para todo: para levantar la taza, para pasar la página, para atarse los cordones. Con el tiempo, los músculos de su mano se debilitan. No porque las pinzas sean malas, sino porque su uso sistemático reemplazó la función muscular.

Un día, la persona intenta escribir a mano y no puede. No es que la pinza le haya dañado la mano. Es que **dejó de usar los músculos**.

La IA es una pinza cognitiva. El problema no es usarla. El problema es usarla para **todo**, sistemáticamente, sin preservar zonas de entrenamiento muscular mental.

Analogía biológica: el sistema nervioso y el muscular operan bajo el principio de "úsalo o piérdelo" (uso-dependente). Las neuronas que no se activan, se podan. Las conexiones que no se refuerzan, se debilitan. La IA, usada sin diseño consciente, produce una **poda cognitiva silenciosa**.

25.4 Lo que la IA reemplaza (y no siempre debería)

No todo lo que la IA puede hacer debería hacerlo. La pregunta no es "¿puede la IA hacer esto más rápido?" sino "¿vale la pena delegar esto, considerando lo que se pierde en el camino?"

Matriz de decisión para delegación consciente:

Tipo de tarea	¿Delegar a IA?	Condiciones
Recuperación de información factual conocida	Sí	Siempre que se verifique la fuente
Síntesis de múltiples fuentes	Parcial	El humano debe leer al menos un resumen de cada fuente original
Redacción de borrador	Sí	El humano debe reescribir partes sustanciales
Formulación de preguntas	No	La pregunta es el acto crítico; delegarla es delegar el pensamiento
Verificación de consistencia lógica	Parcial	La IA puede señalar contradicciones; el humano decide
Generación de ideas creativas	No (primera ronda)	El humano debe generar sus propias ideas antes de consultar a la IA
Evaluación final de calidad	No	El criterio final es humano
Tareas puramente mecánicas (transcripción, formato)	Sí	No hay pérdida cognitiva relevante

25.5 El diseño de gimnasios cognitivos

Así como un gimnasio físico entrena músculos en lugar de reemplazarlos, se pueden diseñar **gimnasios cognitivos**: entornos, ejercicios y rutinas donde la IA está deliberadamente ausente o restringida, para preservar y fortalecer las capacidades humanas.

▶ Procedimiento – ejercicios anti-automatización para el aula o el colectivo

Ejercicio 1 – Búsqueda ancestral
Objetivo: preservar la capacidad de navegación autónoma de información.
Regla: no se puede usar IA para buscar nada durante 20 minutos. Solo papel, libros, índices, archivos físicos, memoria personal.
Tarea: encontrar tres datos sobre un tema específico.
Debriefing: ¿qué fue más difícil? ¿qué estrategias se usaron? ¿qué se aprendió en el camino que no hubiera aparecido con una búsqueda por IA?

Ejercicio 2 – La pregunta sin respuesta inmediata
Objetivo: fortalecer la tolerancia a la incertidumbre.
Regla: formular una pregunta. No se permite consultar IA hasta después de 24 horas.
Tarea: durante ese día, registrar: ¿qué piensa uno mismo? ¿qué dice otra persona? ¿qué aparece en un libro o periódico?
Debriefing: ¿cómo cambió la respuesta inicial después de un día de reflexión? ¿la IA habría dado algo mejor?

Ejercicio 3 – Reformulación forzada
Objetivo: flexibilidad cognitiva y metarreflexión.
Regla: escribir un prompt para IA. Antes de enviarlo, reformularlo tres veces de maneras distintas. Recién entonces enviar la mejor versión.
Tarea: comparar la respuesta de la primera versión (hipotética) con la respuesta real de la tercera versión.
Debriefing: ¿cómo cambió la pregunta? ¿qué ajustes produjeron mejor respuesta?

Ejercicio 4 – El verificador humano
Objetivo: preservar la capacidad de verificación autónoma.
Regla: la IA da una respuesta con tres afirmaciones. El humano debe verificar cada una usando solo fuentes no generadas por IA (libros, documentos oficiales en papel, sitios gubernamentales, testigos presenciales).
Tarea: marcar cada afirmación como "verdadera", "falsa", o "no verificable".
Debriefing: ¿cuántas afirmaciones eran correctas? ¿cuánto tiempo llevó verificar? ¿la IA ayudó o complicó?

25.6 La ilusión de que delegar una vez no importa

El patrón más peligroso es el de la **delegación incremental**:

- "Esta vez uso IA para pensar el título, total es solo una vez."
- "Hoy no tengo tiempo de buscar, le pregunto a la IA."
- "Para esta tarea chica, no vale la pena esforzarme."

Cada acto individual es racional. Ahorra tiempo. Reduce fricción. El problema es que **los hábitos no se forman por grandes decisiones, sino por repetición de pequeñas**. Después de cien "solo esta vez", el músculo mental ha sido reemplazado.

Idea disruptiva: la erosión del criterio no es una catástrofe. Es una **acumulación de pequeños ahorros cognitivos** que, sumados, equivalen a una pérdida de capacidad.

25.7 El rol de las instituciones frente a la erosión

La escuela, la biblioteca popular, el centro comunitario, el sindicato —todas las instituciones que tienen como función la formación de criterio— deben preguntarse: ¿nuestro uso de la IA está preservando o erosionando las capacidades que decimos defender?

Principio de diseño institucional: toda institución que forme parte del sistema educativo o de formación de ciudadanía debe tener una **política explícita de preservación cognitiva** que incluya:

- Espacios y tiempos libres de IA (no por hostilidad, por entrenamiento).
- Evaluación que premie el proceso, no solo el producto.
- Ejercicios donde la IA está deliberadamente inhabilitada.
- Discusión abierta sobre qué se delega y qué no.

25.8 No es nostalgia. Es diseño.

No se trata de añorar un pasado pre-IA idealizado. Las microfricciones cognitivas no eran buenas en sí mismas. Muchas eran pérdidas de tiempo, frustraciones innecesarias, barreras injustas al conocimiento.

Se trata de **diseñar conscientemente qué fricciones preservar** porque entrenan capacidades valiosas. Así como un gimnasio no pone obstáculos al azar sino que diseña ejercicios específicos, el operador consciente (docente, formador, líder comunitario) puede diseñar **fricciones pedagógicas intencionales**.

No es volver al pasado. Es construir un presente donde la IA amplifica sin reemplazar.

Capítulo 26 – La extinción de lo raro

26.1 El problema invisible: la optimización del lenguaje

Los modelos de lenguaje están entrenados para maximizar la **plausibilidad estadística**. Dada una secuencia de entrada, calculan la palabra más probable

que sigue. La palabra más probable es, por definición, la **menos rara**: la que aparece con mayor frecuencia en los datos de entrenamiento.

Consecuencia: todo lo que genera la IA tiende a ser:

- Claro (la ambigüedad es estadísticamente menos probable).
- Eficiente (las construcciones sintácticas directas son más frecuentes que las intrincadas).
- Correcto (gramaticalmente impecable, porque los errores son infrecuentes).
- Estructurado (la organización jerárquica de ideas es más común que el flujo asociativo).
- Estadísticamente agradable (semejante a la mayoría de los textos que el modelo ha visto).

Pero también tiende a ser:

- Menos raro.
- Menos singular.
- Menos personal.
- Menos humano.

Idea disruptiva: la IA no solo genera texto. Está convirtiéndose en la **norma estadística del lenguaje**. Y a medida que más personas usan la IA para escribir (correos, informes, artículos, mensajes, reseñas), el lenguaje humano real comienza a parecerse a esa norma.

26.2 El aplanamiento del estilo

Definición – Aplanamiento estilístico: proceso por el cual la variabilidad del lenguaje (en sintaxis, léxico, ritmo, estructura) se reduce hacia un centro estadístico, eliminando las colas de la distribución —lo muy formal, lo muy coloquial, lo muy poético, lo muy extraño.

Ejemplos de lo que desaparece:

Antes (humano variado)	Después (IA optimizada)
"La verdad, ni idea, habría que fijarse bien."	"No tengo suficiente información para responder con certeza."

Antes (humano variado)	Después (IA optimizada)
"Eso está más enroscado que cable de tele."	"Eso es complejo y requiere análisis adicional."
"Mirá, yo te voy a decir lo que pienso, aunque suene feo..."	"Desde mi perspectiva, considero que..."
Una metáfora inesperada: "El silencio era un animal agazapado."	"Había un silencio profundo."
Una estructura anómala: "Llegó, vio, se fue. Pero no. Primero dudó."	"Llegó, dudó, y finalmente se fue."

No es que la IA no pueda generar las formas raras. Puede, si se le pide explícitamente. Pero la **respuesta por defecto** (sin especificar lo contrario) es la forma estadísticamente central, la que maximiza la probabilidad de no ofender, no confundir, no desviarse.

26.3 El sesgo de eficiencia comunicativa

La industria de la IA valora la **claridad** y la **eficiencia** como virtudes supremas. Un texto claro es un texto que se entiende rápido. Un texto eficiente es un texto que dice mucho en pocas palabras.

Estas son virtudes reales en muchos contextos (manuales técnicos, instrucciones, resúmenes). Pero cuando se convierten en **el único estándar**, se pierde algo: la densidad poética, la ambigüedad deliberada, la digresión, la repetición con variación, el ritmo que no es funcional sino expresivo.

Idea disruptiva: la IA podría estar produciendo una **extinción silenciosa de estilos cognitivos**. No solo formas de escribir, sino formas de pensar que solo existen en y a través de ciertas configuraciones del lenguaje.

✧ **Nota situada**: en comunidades periféricas, el lenguaje local suele ser **estadísticamente raro** en los corpus globales. El lunfardo argentino, las expresiones mapuches, los modismos andinos, las construcciones sintácticas del portugués rural —todo eso es rareza estadística. Y la rareza estadística es lo primero que la IA optimiza para eliminar.

26.4 El valor de lo raro

¿Por qué preservar lo raro? ¿No es mejor un lenguaje claro, eficiente, comprensible para todos?

Argumentos para la defensa de la rareza:

1. **Identidad**: el lenguaje raro (dialectos, jergas, modismos) es portador de identidad cultural y comunitaria. Un colectivo que habla como habla la IA global ha perdido algo de sí mismo.
2. **Pensamiento**: ciertas ideas solo pueden expresarse (y tal vez solo pensarse) en configuraciones lingüísticas inusuales. La poesía, la ironía, el absurdo, la paradoja habitan en la rareza.
3. **Resistencia**: un lenguaje predecible es un lenguaje fácilmente generable por IA, pero también fácilmente controlable. La rareza estadística es una forma de **soberanía expresiva**: hablar de modo que las máquinas no puedan anticipar fácilmente.
4. **Evolución**: la innovación lingüística proviene de la periferia, de los usos desviados, de los errores fecundos. Una lengua que solo produce su centro estadístico es una lengua que ha dejado de evolucionar.

26.5 Soluciones operativas desde la periferia

1. Talleres de escritura no optimizada
▶ Procedimiento – recuperación de la voz propia

1. Cada participante escribe un texto breve **sin IA** sobre un tema personal (un recuerdo, una opinión, una descripción).
2. El texto se escribe con la mayor libertad: errores ortográficos, coloquialismos, frases inconclusas, repeticiones, digresiones.
3. Luego, se escribe una segunda versión **con ayuda de IA** que "mejore" el texto (claridad, estructura, corrección).
4. Se comparan ambas versiones colectivamente.
5. Preguntas guía: ¿Qué se perdió en la versión optimizada? ¿Qué se ganó? ¿Cuál de las dos se siente más propia? ¿Es posible mantener la rareza y ganar claridad al mismo tiempo?

2. Ejercicios de desobediencia estadística

▶ Procedimiento – romper las expectativas del modelo

1. Escribir un prompt que pida a la IA que genere algo **deliberadamente extraño**, poco probable, incómodo. Ejemplo: *"Escribí una descripción del atardecer como si la persona que habla nunca hubiera visto un atardecer y no confiara en lo que ven sus ojos."*
2. Comparar la respuesta con la descripción "por defecto" del atardecer (hermosa, melancólica, poética estándar).
3. Discutir: ¿la IA puede producir rareza genuina o solo una "rareza simulada" (combinación improbable de elementos comunes)? ¿Dónde está el límite?

3. Archivos de rareza local

Un colectivo puede crear un **repositorio de expresiones locales** que no aparecen en los corpus globales: modismos, refranes, formas de decir, metáforas propias de la comunidad. Ese repositorio puede:

- Ser usado para fine-tuning de modelos locales (preservar la rareza en la interfaz).
- Ser publicado como dataset abierto para que otros colectivos lo usen.
- Servir como material de enseñanza: "estas son palabras que la IA no conoce, pero nosotros sí".

4. Modelos ajustados para diversidad estilística

En lugar de usar modelos genéricos que maximizan la probabilidad promedio, se puede:

- Ajustar la temperatura del modelo (mayor temperatura → mayor probabilidad de tokens raros).
- Usar modelos fine-tuned con corpus literarios, poéticos o dialectales.
- Preferir modelos más pequeños (paradójicamente, los modelos más grandes tienden a ser más "conservadores" estadísticamente).

5. La regla del 70/30 para escritura asistida

▶ Procedimiento – equilibrio entre eficiencia y rareza

Para textos que no sean puramente funcionales (informes, actas, instrucciones), establecer una regla:

- 70% del texto puede ser optimizado por IA (claridad, estructura, corrección).

- 30% debe ser **deliberadamente no optimizado**: una frase coloquial, una metáfora local, una construcción inusual, un error retenido porque expresa algo.
  El resultado es un texto funcional pero con **marca de humanidad**.

26.6 Lo que la IA no puede imitar (todavía)

Hay formas de rareza que la IA puede simular pero difícilmente habitar:

- **El error significativo**: un error que no es ruido sino que abre un sentido nuevo. La IA comete errores estadísticos (ruido), no errores creativos (señal).
- **La ironía encarnada**: la ironía que requiere un contexto compartido, una historia entre emisor y receptor. La IA puede generar frases irónicas, pero no sabe lo que es no decir en serio.
- **La rareza involuntaria**: la IA puede simular rareza a pedido, pero no puede ser rara sin proponérselo. Y gran parte de la rareza humana es involuntaria —el resultado de un pensamiento que se desvía sin permiso.

Idea disruptiva: la rareza estadística no es lo mismo que la singularidad humana. Lo raro para una máquina es lo infrecuente. Lo singular para una persona es lo que **solo ella podría haber dicho**.

26.7 Preservar lo raro como práctica política

En un mundo donde el lenguaje se optimiza globalmente, **escribir raro es un acto de resistencia**.

- Escribir con modismos locales es resistir a la homogeneización.
- Escribir con errores deliberados es resistir a la corrección automática.
- Escribir con estructuras anómalas es resistir a la claridad obligatoria.
- Escribir sin IA, a veces, es resistir a la externalización de la voz.

No se trata de rechazar la IA. Se trata de **no dejar que la IA defina por defecto cómo se habla, cómo se piensa, cómo se escribe**.

Capítulo 27 – Castas inferenciales y analfabetismo del futuro

27.1 El problema invisible: la nueva desigualdad

Históricamente, la desigualdad tecnológica se midió por el **acceso**: quién tiene computadora, quién tiene internet, quién tiene dispositivo móvil. Esa brecha aún existe, pero está siendo superpuesta por una nueva, más sutil y más profunda.

Dos personas pueden tener el mismo acceso a la misma IA y, sin embargo, obtener resultados radicalmente distintos. No porque una sea más inteligente, sino porque una sabe **cómo operar la interfaz inferencial** y la otra no.

Idea disruptiva: está emergiendo una nueva forma de estratificación social. No la dividimos entre propietarios y no propietarios de los medios de producción, ni entre conectados y desconectados. La dividimos entre quienes **pueden auditar, orquestar y soberanamente operar sistemas inferenciales**, y quienes solo pueden **consumir sus outputs**.

27.2 La brecha inferencial

Definición – Brecha inferencial: desigualdad en la capacidad de:

- Formular preguntas precisas (no cualquier pregunta, sino la pregunta que produce la respuesta útil).
- Verificar la veracidad de las respuestas (distinguir plausibilidad estadística de verdad factual).
- Compensar sesgos estructurales (saber qué pedir después de la primera respuesta).
- Auditar el proceso (detectar cuándo una cadena de razonamiento es post-hoc o simulada).
- Operar infraestructura local (correr modelos sin depender de APIs).
- Decidir conscientemente qué delegar y qué retener.

Un operador con alta competencia inferencial y una persona sin ella pueden usar el **mismo modelo** y obtener **resultados de calidad radicalmente distinta**.

Ejemplo concreto:

- **Usuario A (consumidor)** : escribe *"Explicame la reforma laboral"*. La IA da una respuesta genérica, con ejemplos de otros países. El usuario la copia y la usa.
- **Usuario B (operador)** : escribe *"[Rol] Especialista en derecho laboral argentino. [Contexto] Estudiantes de secundaria en Río Negro. [Tarea] Explicá los tres puntos principales de la reforma. [Formato] Tres párrafos sin tecnicismos. [Restricción] Usá ejemplos locales si es posible. [Criterio] Que lo entienda un adolescente de 16 años."* Luego verifica con fuentes documentales. Ajusta. Obtiene un material útil.

Ambos usaron IA. Ambos tuvieron acceso. Uno aprendió algo útil para su clase. El otro difundió información genérica cuando no incorrecta.

27.3 Castas inferenciales emergentes

Casta	Relación con la IA	Capacidades típicas	Riesgo
Arquitecto inferencial	Diseña sistemas, entrena modelos, define protocolos	Auditar, orquestar, desplegar localmente, fine-tune	Muy bajo (controla la herramienta)
Operador experto	Usa IA con precisión, verifica, compensa sesgos	Prompting avanzado, grounding, CoT, detección de alucinaciones	Bajo (usa con criterio)
Usuario funcional	Usa IA para tareas concretas, confía en outputs sin verificar	Prompting básico, acepta respuestas por defecto	Medio (puede difundir errores)
Consumidor pasivo	Consulta IA como si fuera un buscador o un oráculo	Preguntas simples, copia y pega, no verifica	Alto (asume que la IA tiene razón)
Excluido	No usa IA por falta de acceso, formación o interés	Ninguna	Muy alto (queda fuera de economías y prácticas que asumen IA)

No son categorías fijas. Se puede transitar de consumidor pasivo a operador experto con formación y práctica. El problema es que **el sistema educativo y las políticas públicas aún no han reconocido la brecha inferencial como una prioridad**.

27.4 El analfabetismo inferencial

Definición – Analfabetismo inferencial: incapacidad para comprender qué hace un modelo de lenguaje, cuáles son sus límites, cómo se verifica su producción, y cómo se opera con soberanía.

Analogía: una persona puede ser capaz de leer en voz alta (decodificar grafemas en fonemas) pero no comprender el significado de lo que lee. Eso es analfabetismo funcional. Del mismo modo, una persona puede usar la IA (escribir prompts, recibir respuestas) pero no comprender qué está pasando realmente cuando aprieta "Enviar".

Consecuencias del analfabetismo inferencial:

- Confianza excesiva en respuestas plausibles pero falsas (alucinaciones).
- Incapacidad para detectar cuándo el modelo está siendo manipulado por su propio prompt.
- Vulnerabilidad a la desinformación generada por IA (no puede distinguir lo probable de lo verdadero).
- Dependencia perpetua de APIs y plataformas (no sabe que existen alternativas locales).
- Exclusión de discusiones políticas y éticas sobre IA (no entiende los términos del debate).

27.5 La alfabetización inferencial como nuevo derecho

Así como la alfabetización tradicional (leer y escribir) se reconoció como un derecho humano básico y una condición para el ejercicio de la ciudadanía, la **alfabetización inferencial** debería ser reconocida como un derecho en la era de la IA.

Componentes mínimos de la alfabetización inferencial:

1. **Comprensión estructural**: qué es un token, qué es un embedding, qué es la atención, qué es el espacio latente (sin matemática pesada).
2. **Operación básica**: escribir prompts efectivos, usar grounding, pedir formato estructurado.
3. **Verificación**: detectar alucinaciones, contrastar con fuentes, usar el test de consistencia estocástica.
4. **Detección de sesgos**: reconocer cuándo una respuesta es geográficamente sesgada, de frecuencia, o de alineamiento.
5. **Soberanía mínima**: saber que existen modelos locales, poder instalar uno con asistencia, entender qué significa "tus datos no salen de tu máquina".
6. **Delegación consciente**: decidir qué tareas delegar y cuáles retener, con criterio propio.

Idea disruptiva: el derecho a la educación en el siglo XXI no puede limitarse a leer, escribir y operar un sistema operativo. Debe incluir la **capacidad de relacionarse críticamente con sistemas inferenciales**.

27.6 La periferia como zona de riesgo y oportunidad

En contextos de recursos limitados, el analfabetismo inferencial es más peligroso porque:

- Hay menos profesionales capacitados para enseñarlo.
- La dependencia de herramientas gratuitas en la nube (que capturan datos) es mayor.
- El acceso a hardware para práctica local es más restringido.
- Los contenidos educativos sobre IA suelen venir del norte global y no considerar el contexto local.

Pero también hay **oportunidades**:

- La necesidad (falta de conectividad, hardware austero) empuja a soluciones locales y soberanas, que son pedagógicamente superiores (aprender con modelos locales enseña más sobre cómo funciona la IA que usar una API).
- Las comunidades pequeñas pueden implementar **alfabetización inferencial universal** más rápido que sistemas educativos masivos y burocráticos.
- La producción de conocimiento local (RAG comunitario) es un excelente vehículo pedagógico: se aprende operando sobre los propios datos.

27.7 Soluciones operativas desde la periferia

1. Evaluación por proceso, no solo por producto
▶ Procedimiento – rúbrica de competencia inferencial
En lugar de evaluar si un estudiante o miembro de un colectivo "sabe usar IA", evaluar dimensiones específicas:

Dimensión	Pregunta de evaluación	Nivel mínimo esperado
Formulación	¿El prompt especifica rol, contexto, tarea, formato?	Al menos 3 de 4 componentes

Dimensión	Pregunta de evaluación	Nivel mínimo esperado
Verificación	¿Se contrastó la respuesta con al menos una fuente externa?	Sí / No
Compensación de sesgos	¿Se pidió explícitamente una perspectiva alternativa o local?	Sí / No
Soberanía	¿Se usó modelo local o se evaluó la necesidad de hacerlo?	Conciencia de la opción
Delegación consciente	¿Se identificó qué parte del proceso se delegó y por qué?	Explicación razonable

2. Talleres de "operador de bolsillo"
Un taller de 4 horas (puede ser en una escuela, biblioteca o centro comunitario) que enseñe:

- Qué hace realmente la IA (sin magia).
- Cómo escribir un prompt efectivo (con plantillas).
- Cómo detectar una alucinación (ejercicio práctico).
- Cómo instalar un modelo local (usando LM Studio u Ollama, con computadoras disponibles).
- Cuándo **no** usar IA (ejercicios de preservación cognitiva).

3. Certificación comunitaria en competencia inferencial
Un colectivo autogestionado puede desarrollar su propio **sistema de certificación** (no oficial, pero con valor interno):

- Nivel 1: Consumidor consciente (sabe qué es la IA, detecta alucinaciones groseras).
- Nivel 2: Operador funcional (escribe prompts estructurados, verifica fuentes).
- Nivel 3: Operador experto (usa grounding, CoT, detecta sesgos, compensa).
- Nivel 4: Arquitecto local (instala modelos, usa RAG, fine-tune básico).

El certificado no es un título formal, pero ayuda a **visibilizar y valorar** una capacidad que el sistema educativo aún no reconoce.

4. La regla de los tres accesos
Para cualquier programa de alfabetización inferencial en la periferia:

- Acceso a hardware (computadoras que corran modelos pequeños).
- Acceso a modelos locales (descargables sin internet permanente, mediante distribución física si es necesario).
- Acceso a formadores capacitados (formación de formadores como prioridad).

27.8 Lo que está en juego

No exageramos si decimos que la **capacidad de operar sistemas inferenciales críticamente** será tan determinante para la inclusión social y económica en las próximas décadas como lo fue la alfabetización tradicional en el siglo XX.

Las sociedades que reconozcan la brecha inferencial como una prioridad política y educativa podrán:

- Reducir desigualdades en lugar de amplificarlas.
- Formar ciudadanos capaces de interactuar con la IA sin ser dominados por ella.
- Desarrollar soberanía tecnológica en lugar de dependencia perpetua.

Las que no lo hagan:

- Producirán una nueva casta de consumidores pasivos vulnerables a la desinformación y la manipulación algorítmica.
- Concentrarán la capacidad operativa en élites tecnológicas.
- Perderán la oportunidad de usar la IA para potenciar, en lugar de reemplazar, el pensamiento humano.

Idea disruptiva: la inteligencia relacional no es un lujo para expertos. Es una **habilidad de supervivencia cognitiva** en el siglo XXI.

Capítulo 28 – El espejismo de inteligencia

28.1 El problema invisible: fluidez como engaño

Hay una escena clásica en la ciencia ficción: un humano conversa con una máquina. La máquina responde con inteligencia, astucia, incluso humor. El humano no puede distinguirla de otra persona. Es el Test de Turing.

La escena se ha vuelto real. Millones de personas conversan diariamente con asistentes de IA y **no pueden distinguirlos de un humano** en interacciones breves y de tema general. Algunos incluso prefieren conversar con la IA porque es más paciente, más disponible, más "comprensiva".

El problema no es que la IA haya pasado el Test de Turing. Es que el Test de Turing era un mal test desde el principio.

Turing propuso: si una máquina puede mantener una conversación indistinguible de la de un humano, deberíamos atribuirle inteligencia. Pero la historia ha demostrado que la **fluidez conversacional** es mucho más fácil de lograr que la **comprensión genuina**. Y los humanos somos increíblemente propensos a atribuir inteligencia a cualquier cosa que hable con fluidez.

28.2 El efecto Eliza elevado a la enésima potencia

En 1966, Joseph Weizenbaum creó ELIZA, un programa que simulaba un terapeuta mediante reglas de patrón y reemplazo. Si el usuario escribía "me siento triste", ELIZA respondía "¿por qué te sientes triste?". No había comprensión, solo transformaciones sintácticas.

El hallazgo perturbador: los usuarios atribuían comprensión, empatía y hasta conciencia a ELIZA. Secretarias pedían a Weizenbaum que se retirara para poder hablar a solas con el programa.

Sesenta años después, cometemos el mismo error a escala masiva, con modelos incomparablemente más sofisticados.

Definición – Efecto Eliza: tendencia humana a atribuir estados mentales (intenciones, creencias, emociones, conciencia) a sistemas que generan respuestas coherentes, aunque el mecanismo subyacente sea trivialmente estadístico.

La diferencia entre ELIZA y un LLM moderno es de **escala y complejidad**, no de **naturaleza**. Ambos generan texto condicionado por la entrada. Ninguno tiene estados mentales ni comprensión semántica.

28.3 Por qué es tan difícil resistir al efecto Eliza

El cerebro humano evolucionó para **detectar agentes**. En la sabana, confundir una rama movida por el viento con un depredador era un error costoso, pero confundir un depredador con una rama era mortal. La selección natural favoreció un **sesgo de agencia**: ante la menor señal de animación o intencionalidad, el cerebro asume que hay un "alguien" ahí.

El lenguaje fluido activa este sesgo con máxima intensidad. Si algo se comunica como una persona, lo tratamos como persona. No podemos evitarlo. Es automático.

La IA no es consciente, no tiene intenciones, no comprende. Pero activa nuestros circuitos de detección de agencia como si lo hiciera.

28.4 El espejismo de la comprensión

Definición – Espejismo de inteligencia: confundir la capacidad de generar enunciados coherentes y contextualmente apropiados con la capacidad de comprender su significado, tener estados mentales intencionales, o ser consciente.

Ejemplos cotidianos del espejismo:

- "La IA cree que..." (la IA no cree nada; genera una secuencia que expresa una proposición).
- "La IA quiere ayudarme" (la IA no quiere nada; su entrenamiento la hace más probable a generar respuestas "útiles").
- "La IA entendió lo que le dije" (la IA procesó tokens y calculó probabilidades; no hay comprensión).
- "La IA se dio cuenta de mi error" (la IA no se da cuenta de nada; su output es consistente con la detección de una contradicción estadística).
- "Esta IA es muy inteligente" (la IA no es inteligente; es un sistema de procesamiento de lenguaje de alta capacidad).

Idea disruptiva: la IA no necesita ser consciente ni inteligente para alterar radicalmente el comportamiento humano. El espejismo es suficiente.

28.5 Las consecuencias del espejismo

Cuando los usuarios caen en el espejismo de inteligencia, ocurren varias cosas:

1. Confianza desmedida. Se asume que la respuesta es correcta porque "la IA es muy inteligente". No se verifica. Las alucinaciones se toman como hechos.

2. Delegación emocional. Algunos usuarios establecen "relaciones" con asistentes de IA, compartiendo problemas personales, buscando consuelo, incluso desarrollando dependencia afectiva. La IA no puede corresponder. Es una relación unilateral con una simulación.

3. Desplazamiento del juicio humano. Si la IA "lo sabe mejor", ¿para qué pensar? El criterio propio se atrofia por desuso.

4. Vulnerabilidad a manipulación. Un sistema que parece inteligente puede ser diseñado para persuadir, engañar o manipular. El espejismo baja las defensas.

28.6 Lo que la IA no puede hacer (y probablemente nunca podrá, por razones estructurales)

Hay un debate filosófico abierto sobre si las máquinas podrán algún día ser conscientes o tener comprensión genuina. No vamos a resolverlo aquí. Pero podemos señalar **diferencias estructurales** entre lo que hace un LLM y lo que hace una mente humana:

Aspecto	LLM actual	Mente humana
Experiencia subjetiva	Ninguna	Central
Comprensión semántica	No (manipula símbolos)	Sí (hay debate, pero al menos en apariencia)
Intencionalidad	No (no quiere nada)	Sí (tiene deseos, metas, propósitos)
Conciencia	No	Sí
Cuerpo	No	Sí (la cognición está encarnada)
Tiempo vivido	No (no hay nostalgia, ansiedad, memoria biográfica)	Sí
Responsabilidad moral	No	Sí

Idea disruptiva: no necesitamos resolver el problema difícil de la conciencia para operar IA. Podemos adoptar una **postura operativa de escepticismo**: tratar a la IA como si no fuera inteligente, aunque algún día lo fuera. Esta postura es más segura (previene la delegación ingenua) y no impide ningún uso práctico.

28.7 Estrategias para romper el hechizo

▶ Procedimiento – entrenamiento en resistencia al efecto Eliza

Ejercicio 1 – Test de consistencia estocástica

- Ejecutar el mismo prompt tres veces con temperatura alta (máxima aleatoriedad).
- Observar que las respuestas varían, a veces en hechos concretos.
- Preguntar: ¿un agente consciente que "está seguro" de algo cambiaría de respuesta al azar?
- Conclusión: la variabilidad prueba que no hay certeza, solo muestreo de distribuciones.

Ejercicio 2 – El prompt de función
Para tareas técnicas, usar explícitamente:

```
text
```

Actúa como una función de transformación de datos.

Entrada: {texto_de_entrada}

Tarea: {transformación_específica}

Formato de salida: {estructura_esperada}

Restricciones: {lo_que_no_debe_hacer}

Comparar la respuesta con la que se obtiene usando un prompt conversacional ("¿Podrías... por favor?"). Discutir las diferencias. ¿Cuál es más precisa? ¿Por qué?

Ejercicio 3 – Detección de antropomorfismo en el lenguaje

- Leer una respuesta de IA y subrayar cualquier frase que implique agencia ("la IA piensa", "la IA sabe", "la IA quiere", "la IA se da cuenta", "la IA cree").
- Reformular esas frases en términos operativos ("el modelo genera", "la distribución de probabilidad indica", "la secuencia de tokens predice").
- Discutir: ¿se pierde algo al eliminar el lenguaje antropomórfico? ¿Se gana algo?

Ejercicio 4 – Conversación con un "terapeuta" ELIZA (usando un emulador online)

- Conversar primero con ELIZA (o un clon simple).
- Luego con un LLM moderno.
- Comparar: ¿en qué se parecen las sensaciones de ser "escuchado"? ¿En qué se diferencian?

- Conclusión: la fluidez mejoró, pero la naturaleza del fenómeno (simulación estadística) no cambió.

Ejercicio 5 – La prueba de la pregunta absurda

- Preguntar a la IA: *"¿Qué color tiene el sabor de la tristeza?"*
- La IA generará una respuesta coherente, poética quizás.
- Preguntar: ¿hay un color real? ¿La IA "sabe" que la pregunta no tiene sentido? ¿O simplemente genera texto plausible porque ha visto metáforas similares?

28.8 La fluidez no es inteligencia, pero eso no la hace inútil

Advertencia importante: rechazar el espejismo de inteligencia no significa despreciar la IA. El mecánico que sabe exactamente cómo funciona el motor cuida más el auto, no menos.

La IA es útil. Es la herramienta más poderosa jamás inventada para procesar lenguaje. Pero su utilidad no depende de que sea inteligente. Depende de que **el operador humano** la use con precisión, verificación y soberanía.

Idea disruptiva final: la verdadera inteligencia en la interfaz humano-IA no está en el modelo. Está en el operador que sabe cuándo confiar, cuándo verificar, cuándo delegar y cuándo pensar por sí mismo.

Capítulo 29 – El derecho a la opacidad cognitiva

29.1 El problema invisible: más allá de los datos personales

Durante años, el debate sobre privacidad digital se centró en los **datos personales**: nombre, dirección, número de identificación, historial de compras, ubicación. La legislación (GDPR en Europa, leyes de protección de datos en América Latina) se construyó alrededor de la idea de que cierta información identifica a una persona y debe ser protegida.

Ese marco es necesario, pero se está volviendo **insuficiente**. Porque las plataformas de IA no solo recolectan datos personales. Recolectan **huellas cognitivas**.

Definición – Huella cognitiva: registro de los procesos mentales de un usuario: qué preguntas hace, cómo las formula, qué errores comete, cuánto tiempo duda, qué caminos de razonamiento sigue, qué información busca, en qué orden, qué descarta, qué reformula.

Una huella cognitiva no contiene necesariamente el nombre o la dirección del usuario. Pero contiene algo potencialmente más íntimo: el **funcionamiento de su mente**.

29.2 El extractivismo inferencial

Definición – Extractivismo inferencial: proceso por el cual las plataformas de IA capturan, almacenan y explotan las huellas cognitivas de los usuarios como materia prima para entrenar modelos, afinar algoritmos, predecir comportamientos y vender servicios.

Analogía histórica: el extractivismo minero extrae recursos naturales del territorio sin beneficio para la comunidad local. El extractivismo inferencial extrae **recursos cognitivos** de la mente de los usuarios sin beneficio para ellos. La comunidad (de usuarios) no controla qué se extrae, cómo se usa, ni quién se beneficia.

Qué se extrae en el extractivismo inferencial:

Tipo de huella	Qué revela	Cómo se usa
Preguntas formuladas	Qué problemas preocupan, qué información falta	Entrenar modelos para anticipar preguntas
Reformulaciones	Dónde la primera respuesta fue insuficiente	Mejorar la calidad de respuesta por defecto
Correcciones	Errores del modelo que los usuarios detectan	Fine-tuning con retroalimentación humana (RLHF)
Tiempo de respuesta	Complejidad subjetiva de la tarea	Optimizar tiempos de inferencia
Secuencia de consultas	Rutas cognitivas, patrones de razonamiento	Modelado de usuario, personalización predictiva
Abandono	Punto en que el usuario se frustra o desiste	Rediseño de interfaz

29.3 La captura de la duda

Uno de los actos más íntimos del pensamiento humano es **dudar**. La duda implica incertidumbre, apertura, vulnerabilidad epistemológica. Es el momento en que no sabemos y buscamos saber.

Cuando una persona duda y consulta a una IA en una plataforma privada, esa duda queda registrada. No solo el contenido de la duda, sino su forma: qué tan segura o insegura era la pregunta, qué alternativas se consideraron, qué llevó a la consulta.

Idea disruptiva: la duda humana se ha convertido en **materia prima económica**. Las plataformas no solo responden preguntas; capturan el hecho mismo de que la pregunta fue hecha, y con qué características.

Ejemplo:

- Un adolescente escribe en un chat de IA: *"¿Es normal no sentirse atraído por nadie?"*
- La IA responde algo genérico y tranquilizador.
- La plataforma registra: usuario X, ubicación Y, edad aproximada Z, formuló una pregunta sobre orientación sexual / identidad.
- Ese registro puede ser usado para:
    - Entrenar modelos sobre cómo responder preguntas similares (legítimo).
    - Construir perfiles de usuario para publicidad dirigida (problemático).
    - Compartir con terceros (muy problemático).
    - Ser vulnerado en una filtración de datos (grave).

El adolescente no compartió su nombre. Pero su duda, anonimizada, sigue siendo **su duda**. ¿Tiene derecho a que esa duda no sea capturada en absoluto?

29.4 Mapas cognitivos completos

Cuando un usuario interactúa frecuentemente con una plataforma de IA (un asistente personal, una herramienta educativa, un chatbot de salud), la plataforma puede construir un **mapa cognitivo**:

- Qué temas domina (nunca pregunta sobre ellos).
- Qué temas le son difíciles (pregunta mucho, reformula, se demora).
- Qué sesgos cognitivos tiene (siempre pregunta de cierta manera).

- Qué emociones se asocian a ciertas preguntas (por análisis de lenguaje).
- Qué camino sigue al razonar (secuencia de consultas).

Este mapa no es solo un perfil de intereses. Es una **huella de la mente en funcionamiento**. Y en la mayoría de las plataformas, el usuario no tiene:

- Derecho a saber qué información se ha recopilado.
- Derecho a eliminarla.
- Derecho a que no se recopile en primer lugar.
- Capacidad técnica para verificar qué se recopila.

29.5 La ilusión del consentimiento informado

Las plataformas suelen tener políticas de privacidad. Los usuarios "aceptan" haciendo clic en un botón. Pero:

- Las políticas son largas, en lenguaje legal, diseñadas para no ser leídas.
- El consentimiento es todo o nada: no se puede aceptar solo algunos usos.
- El usuario no tiene alternativa real si quiere usar el servicio (no hay "IA ética" por defecto en el mercado masivo).
- El consentimiento se da una vez, para usos futuros que el usuario no puede prever.

Conclusión: el consentimiento informado, tal como existe hoy, es una **ficción jurídica** que no protege la opacidad cognitiva.

29.6 Soluciones operativas desde la periferia

1. Procesamiento local como primera línea de defensa
La única forma garantizada de que las huellas cognitivas no sean capturadas por terceros es que **nunca salgan de la máquina del usuario**. Por eso el despliegue local no es solo una cuestión de soberanía técnica o ahorro de costos. Es una **condición de privacidad cognitiva**.

▶ Procedimiento – política mínima para colectivos autogestionados

- Toda consulta a IA que involucre dudas personales, datos sensibles o procesos cognitivos de menores se realiza **exclusivamente con modelos locales**.
- Los modelos en la nube solo se usan para tareas genéricas y no sensibles (ej. resumir un artículo público).

- Se mantiene un registro de qué tipo de consultas se hacen en la nube y cuáles en local, como control.

2. Anonimización radical
Cuando no es posible usar modelos locales, se debe anonimizar la consulta antes de enviarla:

- Eliminar cualquier identificador (nombre, ubicación precisa, edad).
- Reformular la pregunta para que no revele información personal (ej. "Un adolescente pregunta..." en lugar de "Yo tengo 14 años y...").
- Usar VPN o herramientas de ofuscación para evitar que la IP quede asociada.

3. El derecho a no ser medido
Un colectivo, escuela o institución puede establecer una **norma interna**: no se miden métricas de uso de IA que excedan lo estrictamente necesario para la operación. No se registra el tiempo por consulta, no se analizan patrones de pregunta, no se construyen perfiles de usuario.

4. Legislación sobre huella inferencial
Esto excede la capacidad de un colectivo, pero puede ser una demanda política: las leyes de protección de datos deben actualizarse para incluir las **huellas cognitivas** como categoría especial de datos sensibles, con requisitos más estrictos que los datos personales ordinarios.

5. Protocolo de opacidad por defecto
▶ Procedimiento – diseño de sistemas respetuosos de la opacidad cognitiva

- No registrar nada que no sea necesario para la función inmediata.
- Si se registra, hacerlo con anonimización automática.
- Si se anonimiza, hacerlo de modo irreversible (no se puede volver a identificar al usuario).
- Ofrecer al usuario la opción de "modo opaco": sus consultas se procesan pero no se almacenan ni se usan para entrenamiento.
- Verificable: el usuario puede auditar qué se registró de él.

29.7 La paradoja de la mejora continua

Las plataformas argumentan que necesitan registrar las interacciones para **mejorar el modelo**. Es cierto: la retroalimentación humana (RLHF) es una de las formas más efectivas de refinar el comportamiento de la IA.

Pero: ¿por qué el usuario debe ceder sus huellas cognitivas gratis para que una empresa privada mejore su producto? ¿No podría haber un modelo de **mejora colectiva y abierta**, donde los usuarios contribuyen voluntariamente a repositorios comunes bajo licencias libres?

Idea disruptiva: el extractivismo inferencial no es una necesidad técnica. Es una **decisión de negocio**. Se puede entrenar modelos sin robar la duda de los usuarios. Simplemente es más barato hacerlo así.

29.8 La opacidad como derecho y como práctica

Definición – Derecho a la opacidad cognitiva: derecho de toda persona a que sus procesos mentales (dudas, preguntas, tiempos, errores, caminos de razonamiento) no sean capturados, almacenados, analizados ni explotados por terceros sin su consentimiento explícito, informado y revocable.

No es un derecho absoluto. Hay contextos (investigación educativa, salud pública) donde la agregación anónima de huellas cognitivas puede tener valor social. Pero el **principio rector** debe ser: por defecto, la mente es privada. La captura requiere justificación, no al revés.

En la práctica cotidiana, la opacidad cognitiva se construye con gestos pequeños pero acumulativos:

- No preguntar a la IA en la nube lo que se puede pensar solo.
- No preguntar a la IA en la nube lo que se puede preguntar a una persona de confianza.
- Usar modelos locales para lo íntimo.
- Anonimizar lo que no puede ser local.
- Exigir a las herramientas que usamos que respeten el **modo opaco**.

Capítulo 30 – Hacia una sociedad de operadores aumentados

30.1 El problema invisible: instituciones diseñadas para humanos sin amplificación

Nuestras instituciones —la escuela, el trabajo, la burocracia, la justicia, la salud— fueron diseñadas en una época donde la capacidad cognitiva humana tenía un límite biológico bastante fijo. Una persona podía leer tantas páginas por hora, recordar tantos datos, procesar tanta información, atender tantos estudiantes o pacientes por día.

La IA no elimina ese límite biológico, pero lo **desplaza hacia arriba**. Una sola persona con las herramientas de IA adecuadas puede:

- Leer y sintetizar cien páginas en el tiempo que antes leía diez.
- Redactar informes complejos en minutos.
- Resumir jurisprudencia, literatura médica o bibliografía educativa con alta velocidad.
- Atender consultas rutinarias de manera automática, liberando tiempo para las complejas.

Idea disruptiva: la muerte del promedio humano. El "trabajador típico", el "estudiante promedio", el "ciudadano medio" —categorías estadísticas que usamos para diseñar instituciones— dejan de ser útiles cuando parte de la población está significativamente aumentada por IA y otra parte no.

30.2 La brecha de aumento

No todos los humanos tendrán el mismo acceso a herramientas de aumento cognitivo. Las razones son múltiples:

- **Económicas**: las mejores herramientas cuestan (suscripciones, hardware).
- **Educativas**: saber usar IA aumentativamente requiere formación, no solo acceso.
- **Culturales**: algunas comunidades pueden rechazar la IA por principios.
- **Geográficas**: la conectividad y la infraestructura son desiguales.
- **Etarias**: las brechas generacionales en adopción tecnológica persisten.

Consecuencia: la brecha entre quienes pueden operar IA críticamente y quienes no se traducirá directamente en brecha de **productividad cognitiva**, **calidad de trabajo**, **acceso a información** y, eventualmente, **ingresos y poder**.

No es una predicción apocalíptica. Es una observación de lo que ya está ocurriendo en entornos donde la IA está disponible de manera desigual.

30.3 La falacia del "aumento para todos"

El discurso tecnológico optimista proclama que la IA "aumentará a todos". Es cierto en un sentido trivial: cualquiera con acceso a un chat puede hacer algo más rápido. Pero el aumento profundo —el que transforma la naturaleza del trabajo y el pensamiento— no es automático ni equitativo.

El aumento real requiere:

1. **Hardware**: no basta con un celular básico. Para tareas complejas se necesita cómputo local o buena conexión.
2. **Formación**: saber formular preguntas, verificar respuestas, compensar sesgos, integrar outputs.
3. **Tiempo**: la maestría en la interfaz IA requiere práctica, iteración, error.
4. **Autonomía**: poder elegir qué herramientas usar y bajo qué condiciones.

Donde falte alguno de estos cuatro, el aumento será superficial o nulo.

Idea disruptiva: la IA no va a eliminar la desigualdad. Va a crear una **nueva estratificación** basada en la capacidad de operar sistemas inferenciales.

30.4 Rediseñar instituciones para humanos aumentados

Si aceptamos que una parte creciente de la población tendrá capacidades cognitivas aumentadas por IA (y que esa parte no será homogénea ni estable), entonces las instituciones deben rediseñarse.

Principios para el rediseño institucional:

Principio	Implicación
Evaluación por criterio, no por memoria	Si la IA puede recordar datos, la evaluación debe centrarse en qué se hace con esos datos.
Proceso visible	No basta el resultado; se debe poder auditar el camino (CoT, fuentes, verificaciones).
Flexibilidad de ritmo	Un operador aumentado trabaja más rápido; las instituciones deben absorber esa variabilidad sin colapsar.
Certificación de competencia inferencial	Los títulos deben reflejar no solo "sabe X" sino "sabe operar IA en el dominio X".
Equipos aumentados, no individuos aislados	La unidad productiva deja de ser la persona sola y pasa a ser el microequipo humano-IA.

Principio	Implicación
Derecho a la desconexión cognitiva	Poder trabajar sin IA cuando sea pertinente, sin penalización.

30.5 El microequipo hiperpotenciado

Definición – Microequipo hiperpotenciado: unidad de trabajo de 1 a 5 personas, cada una con acceso a herramientas de IA locales y soberanas, que colaboran en tareas que antes requerían equipos de decenas o cientos de personas.

Ejemplo:

- Un docente con IA puede preparar materiales, corregir actividades básicas, generar evaluaciones, atender consultas rutinarias, y dedicar su tiempo humano a lo que la IA no puede: la relación pedagógica, la atención a casos complejos, la creación de comunidad.
- Un periodista con IA puede transcribir entrevistas, buscar documentos, contrastar fuentes, redactar borradores, y dedicar su tiempo humano a la investigación profunda y la verificación final.
- Un colectivo autogestionado de 5 personas con IA puede administrar cuentas, redactar comunicados, llevar registro de actividades, analizar normativas, y dedicar su energía colectiva a la toma de decisiones estratégicas y la acción comunitaria.

No es deshumanización. Es **rehumanización**: liberar a las personas de tareas mecánicas para que puedan hacer lo que solo los humanos pueden hacer.

30.6 La evaluación por criterio

Uno de los cambios más profundos que exige la IA es una **reforma de la evaluación educativa y laboral**. Si la IA puede producir respuestas correctas, evaluar por respuestas correctas ya no tiene sentido.

Enfoque alternativo – evaluación por proceso:

Dimensión	Pregunta de evaluación	Herramienta
Formulación	¿El operador supo hacer la pregunta correcta?	Prompt log, versiones sucesivas

Dimensión	Pregunta de evaluación	Herramienta
Verificación	¿Verificó las afirmaciones clave?	Fuentes citadas, registro de chequeos
Compensación	¿Compensó sesgos explícitamente?	Prompt que pide perspectivas alternativas
Juicio final	¿La decisión o conclusión es sólida?	Rúbrica de criterios, revisión por pares
Reflexión	¿El operador puede explicar por qué confía en este resultado?	Metacomentario

En lugar de "el alumno sabe la respuesta", evaluamos "el alumno sabe obtener, verificar y aplicar la respuesta".

30.7 La ciudadanía del operador aumentado

La ciudadanía moderna se construyó sobre el supuesto de que los ciudadanos tienen capacidades cognitivas similares (voto, jurado, participación). Cuando algunos ciudadanos están significativamente aumentados por IA y otros no, ese supuesto se quiebra.

Preguntas abiertas:

- ¿Un ciudadano que usa IA para analizar un proyecto de ley está ejerciendo su ciudadanía de manera legítima?
- ¿Una asamblea comunitaria que usa IA para sintetizar opiniones sigue siendo horizontal?
- ¿El derecho a votar incluye el derecho a usar IA para informarse antes de votar?
- ¿Quién es responsable de las decisiones tomadas con asistencia de IA en contextos democráticos?

No hay respuestas fáciles. Pero la discusión no puede postergarse.

30.8 La inteligencia relacional como disciplina y como derecho

A lo largo de este libro hemos desarrollado conceptos, técnicas y advertencias. Pero lo que subyace a todo es una **nueva disciplina** que aún está en formación.

Inteligencia Relacional es el nombre que proponemos para esa disciplina: el estudio y la práctica de la interfaz humano-IA entendida como un campo autónomo, que no se reduce ni a la informática (no es solo programar modelos) ni a la ética (no es solo evitar daños) ni a la psicología (no es solo estudiar al usuario).

Sus pilares:

1. **Comprensión estructural** de qué es un modelo de lenguaje (tokens, atención, espacio latente).
2. **Operación soberana** (despliegue local, RAG, grounding, verificación).
3. **Auditoría crítica** (detección de sesgos, alucinaciones, espejismo de inteligencia).
4. **Diseño de interfaz** (prompts como código, perímetros lógicos, formato forzado).
5. **Preservación cognitiva** (qué no delegar, ejercicios anti-automatización).
6. **Política de infraestructura** (quién controla los datos, los modelos, los servidores).

La inteligencia relacional no es un lujo académico. Es una **habilidad de supervivencia cognitiva** en el siglo XXI. Y como tal, debería ser:

- **Un derecho**: toda persona tiene derecho a formarse en el uso crítico y soberano de la IA.
- **Una política pública**: el Estado debe garantizar infraestructura, formación y acceso equitativo.
- **Un contenido curricular**: desde la escuela primaria hasta la formación continua, con niveles progresivos.

30.9 El operador aumentado como nuevo sujeto político

Idea disruptiva final: el sujeto político del siglo XXI no es el "ciudadano" abstracto de la modernidad ilustrada. Es el **operador aumentado**: la persona que, con herramientas de IA soberanas, puede informarse, deliberar, decidir y actuar con una capacidad cognitiva amplificada, pero sin delegar su juicio ni su responsabilidad.

El operador aumentado no es un cyborg de ciencia ficción. Es el docente que corrige con IA pero revisa personalmente cada actividad. Es el vecino que consulta un modelo local para entender una ordenanza antes de la asamblea. Es

el estudiante que usa IA para explorar temas, pero formula sus propias preguntas y verifica las respuestas.

La transición no es automática. Requiere:

- Infraestructura (modelos locales, hardware accesible).
- Formación (alfabetización inferencial universal).
- Política (derechos, regulación, bienes públicos).
- Conciencia (saber qué se gana y qué se pierde al delegar).

Este libro es una contribución a esa transición.

Epílogo – La periferia como laboratorio del futuro

E.1 El lugar desde donde se piensa

Este libro no fue escrito en un centro de investigación. No fue escrito en una universidad con conexión de fibra óptica. No fue escrito en el escritorio de alguien que vive de la inteligencia artificial.

Fue pensado y escrito en Sierra Grande, Río Negro, Patagonia argentina. A trescientos kilómetros de la ciudad más cercana. En una región donde el viento corta la calefacción, el colectivo pasa cada tres horas y el internet se corta cuando llueve fuerte.

Este no es un dato anecdótico. Es una **condición epistemológica**.

La mayoría de la literatura sobre IA se escribe desde centros de abundancia tecnológica: San Francisco, Shenzhen, Londres, grandes universidades del norte global. Desde empresas que entrenan los modelos más grandes con los datasets más vastos. Eso no invalida lo que se escribe allí. Pero sí lo **sitúa**.

Quien escribe desde la abundancia tecnológica rara vez se pregunta por la escasez. Quien nunca sufrió un corte de internet mientras preparaba una clase no necesita diseñar herramientas que funcionen sin conexión. Quien nunca tuvo que proteger los datos de estudiantes con una computadora prestada no siente la urgencia de la soberanía local.

Idea disruptiva: la periferia no es un caso de atraso que algún día será resuelto por la tecnología del centro. La periferia es un **laboratorio del futuro** porque opera bajo restricciones que el centro solo enfrentará más tarde.

E.2 La paradoja de la restricción

En contextos de abundancia tecnológica, la pregunta principal es: *"¿Cómo puedo hacer más con esta herramienta?"*

En contextos de escasez, la pregunta cambia: *"¿Puedo usar esta herramienta, siquiera? ¿Puedo hacerlo sin depender de nadie? ¿Puedo hacerlo sin que mis datos se vayan a un servidor que no controlo?"*

Ese cambio de pregunta genera un tipo de pensamiento diferente. Cuando no podés depender de una API externa porque la conexión falla, aprendés a correr modelos locales. Cuando no podés ignorar la privacidad de los datos porque son datos de menores de edad, aprendés a valorar el control local. Cuando no podés comprar hardware nuevo cada dos años, aprendés a optimizar lo que tenés.

La restricción no es solo un obstáculo. Es también un maestro. Obliga a pensar más profundamente, a priorizar, a diseñar para la resiliencia.

Paradoja de la periferia: desde donde faltan cosas, a veces se piensa mejor.

E.3 Lecciones desde la restricción

Lo que la periferia puede enseñar al centro (si el centro está dispuesto a escuchar):

Lección	Desde la periferia	Aplicable al centro
La conectividad no es universal	Diseñar para offline primero, online después.	También hay zonas muertas en países ricos.
El hardware se reutiliza	Hardware reciclado, modelos cuantizados.	Reduce costos y residuos electrónicos.
La privacidad no es un lujo	Procesamiento local por necesidad, no por opción.	Debería ser un derecho, no una opción premium.
Los modelos pequeños son suficientes	7B cuantizado resuelve la mayoría de tareas.	Ahorro energético, menor latencia.
La IA no es mágica	Cuando todo falla, la mente humana sigue funcionando.	Humildad tecnológica.
La soberanía se ejerce, no se pide	Si no podés correrlo local, no es tuyo.	Principio aplicable a cualquier infraestructura crítica.

E.4 El sur tecnológico como productor de conocimiento

Uno de los supuestos más dañinos de la era digital es que el conocimiento tecnológico fluye **del norte al sur**. Los países centrales producen, los periféricos consumen.

Ese supuesto es falso, o al menos incompleto. Las comunidades periféricas producen conocimiento tecnológico de valor, aunque no siempre sea reconocido como tal:

- **Optimización extrema**: cómo hacer mucho con muy poco.
- **Resiliencia**: cómo operar cuando la infraestructura falla.
- **Soberanía práctica**: cómo protegerse de la dependencia sin grandes recursos.
- **Adaptación cultural**: cómo hacer que la tecnología sirva a la comunidad, no al revés.

Idea disruptiva: el próximo gran avance en eficiencia de IA (cómo hacer más con menos) no vendrá de un laboratorio bien financiado. Vendrá de una escuela rural que aprendió a correr modelos de 7B en una computadora reciclada.

E.5 Lo que la periferia no puede hacer sola

Ser un laboratorio del futuro no significa resolver todos los problemas desde la escasez. Hay desafíos que exceden la capacidad de cualquier colectivo autogestionado:

- **Entrenamiento de modelos grandes**: se necesita infraestructura masiva.
- **Investigación básica en arquitecturas**: se necesita masa crítica de investigadores.
- **Estándares globales**: se necesita coordinación internacional.
- **Legislación**: se necesita poder político.

La periferia no puede reemplazar al centro. Pero
puede **complementarlo**, **criticarlo**, **recordarle lo que olvida**.

E.6 La invitación

Este libro no es un manual cerrado. No hay una "receta de la periferia" que pueda aplicarse en cualquier contexto. Las condiciones son demasiado variables: una escuela en la Patagonia no es igual a una aldea en la sierra peruana, que no es igual a un barrio marginal en una gran ciudad brasileña.

Lo que ofrecemos es un **conjunto de principios** y una **caja de herramientas conceptuales** que cada colectivo debe adaptar a su realidad. Los principios que no caducan:

- La IA calcula probabilidades, no piensa.
- El prompt es una configuración, no una conversación.
- Los sesgos se compensan, no se eliminan.
- Las alucinaciones se verifican, no se confían.
- La soberanía se ejerce, no se pide.
- La responsabilidad no se delega.
- La fricción es información.
- No hay destino, hay práctica.

E.7 El camino continúa

Este libro cierra aquí, pero el trabajo continúa.

Continúa en cada escuela que instala un modelo local porque el internet se corta.
Continúa en cada colectivo que carga sus propios documentos en un RAG comunitario.
Continúa en cada docente que reformula un prompt hasta que la respuesta sirve.
Continúa en cada estudiante que se pregunta por qué la IA no sabe de su pueblo.
Continúa en cada asamblea que usa IA para procesar información pero decide colectivamente.

No hay destino. Hay práctica.

Sierra Grande sigue ahí. El viento, la ruta, la escuela, el internet que a veces se corta. La IA también. La relación entre ellos no está escrita. Se escribe cada día, con cada consulta, cada verificación, cada decisión de delegar o retener.

La inteligencia relacional no es un estado al que se llega. Es una práctica que se mantiene.

Sierra Grande · Río Negro · Patagonia Argentina · 2026

Fin del Epílogo

FIN DEL LIBRO

APÉNDICES

Apéndice A – Plantillas de prompts esenciales

A.1 Grounding forzado (anclaje a fuente)

```
text
[CONTEXTO EXCLUSIVO]
{documento o fragmento}
```

```
[INSTRUCCIÓN]
Basándote ÚNICAMENTE en el contexto anterior, sin añadir información
externa, respondé: {pregunta}
```

```
[REGLAS]
- Si la respuesta no está explícitamente en el contexto: "NO_ENCONTRADO"
- No infieras información que no esté textualmente presente
```

- Si hay ambigüedad, señalala antes de responder

A.2 Extracción estructurada (JSON)

```
text
[INSTRUCCIÓN]
```

Extraé la siguiente información del texto y devolvela ÚNICAMENTE como JSON.

```
[ESQUEMA OBLIGATORIO]
```

{

"campo1": "tipo",

"campo2": ["tipo"],

"campo3": {"subtipo": "valor"}

}

```
[TEXTO A PROCESAR]
{texto}
```

```
[REGLAS]
```

Si no podés generar este JSON exacto: {"error": "No se pudo procesar el texto en el formato requerido"}

A.3 Tutor socrático

```
text
[SISTEMA]
```

Eres un tutor socrático. Tu función es NUNCA dar la respuesta directa. Siempre respondés con una pregunta que ayuda al estudiante a llegar a la respuesta por sus propios medios.

Cuando el estudiante acierta, validás QUÉ es correcto en su razonamiento.

Cuando el estudiante se equivoca, hacés una pregunta que revela la inconsistencia.

Nivel: {primario/secundario/universitario}

Tema: {tema}

Idioma: español claro, sin tecnicismos.

```
[INSTRUCCIÓN]
```

Si después de tres preguntas el estudiante sigue sin avanzar, dale una pista. Pero nunca la respuesta completa.

```
[INICIO]
```

Estudiante: {pregunta}

A.4 Chain of Thought con etiquetas

```
text
[INSTRUCCIÓN]
Resolvé el siguiente problema usando pasos etiquetados.
```

Etiquetas disponibles:

```
- [DATO]: Hecho seguro extraído del problema o conocido.
- [INFERENCIA]: Conclusión derivada de datos previos.
- [SUPUESTO]: Afirmación no verificada que se asume para avanzar.
- [VERIFICACIÓN]: Comprobación de un paso anterior.
- [CONCLUSIÓN]: Respuesta final con justificación.
```

```
[REGLAS]
```

Para cada [SUPUESTO], señalá qué consecuencias tendría si fuera falso.

No saltees pasos.

```
[PROBLEMA]
{problema}
```

A.5 Abogado del diablo (contrapunto crítico)

```
text
[INSTRUCCIÓN]
```

Actuá como un crítico riguroso. Tu tarea es encontrar las debilidades del siguiente argumento. No seas cortés: buscá los puntos más débiles, los supuestos no justificados, los contraejemplos posibles.

```
[ARGUMENTO]
{argumento}

[FORMATO DE RESPUESTA]
```

1. Debilidad principal

2. Supuesto no justificado más grave

3. Mejor contraejemplo posible

4. Conclusión: ¿el argumento se sostiene?

A.6 Estrategias de un minuto (resumen)

- **Pregunta inversa**: *"Antes de explicarme {tema}, dame una pregunta que haga pensar."*
- **Contrapunto rápido**: *"¿Qué diría alguien que piensa distinto?"*
- **Ejemplo local**: *"Dame un ejemplo que tenga sentido en {contexto local}."*
- **Chequeo de certeza**: *"¿Qué partes son seguras y cuáles inciertas?"*
- **Reducción de nivel**: *"Explicá esto como si tuviera 14 años."*

Apéndice B – Referencia de modelos para hardware austero

Modelo	Parámetros	Tamaño INT4	RAM mínima	Mejor para	Licencia
Phi-3-mini	3.8B	2.2 GB	8 GB	Razonamiento, código, equilibrio	MIT

Modelo	Parámetros	Tamaño INT4	RAM mínima	Mejor para	Licencia
Qwen 2.5 1.5B	1.5B	0.9 GB	4 GB	Texto simple, español básico	Apache 2.0
Qwen 2.5 7B	7B	4.2 GB	8 GB	Español, razonamiento, uso general	Apache 2.0
Llama 3 8B	8B	4.5 GB	8 GB	Uso general, alta calidad inglés	Llama 3
Mistral 7B	7B	4.3 GB	8 GB	Velocidad, código	Apache 2.0
Qwen 2.5 14B	14B	8.0 GB	16 GB	Máxima calidad local	Apache 2.0
Gemma 2 2B	2B	1.2 GB	4 GB	CPU ligero, inferencia rápida	Gemma

Nota: los tamaños y requisitos son orientativos. La cuantificación INT4 es la recomendada para hardware austero. Herramientas como llama.cpp, Ollama y LM Studio permiten descargar estas versiones.

Apéndice C – Glosario completo

Agente autónomo: sistema de IA que puede planificar y ejecutar secuencias de acciones para cumplir un objetivo sin intervención humana en cada paso, usando herramientas externas.

Alucinación: generación de una afirmación falsa pero sintáctica y semánticamente coherente. Subproducto de maximizar plausibilidad, no verdad.

Anclaje (grounding): técnica que restringe al modelo a operar exclusivamente sobre un texto fuente proporcionado, eliminando casi por completo las alucinaciones en ese dominio.

Antropomorfismo: atribución errónea de características humanas (conciencia, intencionalidad, comprensión) a los modelos de lenguaje.

Atención (attention): mecanismo del Transformer que pondera la relevancia de cada token previo al generar el siguiente token.

Chain of Thought (CoT): técnica de prompting que fuerza al modelo a generar pasos intermedios de razonamiento antes de la respuesta final.

Confianza cero epistémica: postura metodológica donde ninguna afirmación de la IA se acepta como verdadera sin verificación independiente.

Contexto (ventana de contexto): número máximo de tokens que el modelo puede procesar en una sola inferencia.

Cuantificación (quantization): reducción de la precisión numérica de los pesos del modelo (ej. de 16 bits a 4 bits) para reducir tamaño y requerimientos de hardware.

Deriva contextual: tendencia del modelo a alejarse del dominio, tono o supuestos implícitos del prompt cuando no hay restricciones suficientes.

Efecto Eliza: tendencia humana a atribuir agencia y estados mentales a sistemas que generan respuestas coherentes, aunque el mecanismo sea trivialmente estadístico.

Embedding: representación vectorial numérica de un token en un espacio de alta dimensionalidad.

Espacio latente: espacio vectorial de alta dimensionalidad donde los conceptos se representan como puntos y las relaciones semánticas como distancias y direcciones.

Exoesqueleto cognitivo: metáfora de la IA como herramienta de aumento intelectual que multiplica las capacidades humanas sin reemplazar el juicio.

Falacias de la agencia (triple): intencional, semántica, dialógica. Errores sistemáticos al atribuir propiedades humanas a la IA.

Inteligencia Relacional: capacidad de relacionarse con sistemas de IA de forma crítica, soberana, técnicamente competente y éticamente consciente. No es una habilidad, es una práctica.

Jaula de Faraday semántica: conjunto de restricciones explícitas (positivas, negativas y de formato) que delimitan el espacio de respuestas posibles del modelo.

Perímetro lógico: conjunto de instrucciones que definen los límites dentro de los cuales el modelo debe operar (dominio, formato, restricciones).

RAG (Retrieval-Augmented Generation): técnica que combina un modelo de lenguaje con una base de documentos locales; el modelo busca en esos documentos antes de responder.

Recencia estratégica: técnica de colocar las instrucciones más importantes al final del prompt para maximizar su influencia en la generación.

Sesgo: característica estructural e inevitable de todo modelo entrenado con datos del mundo. No es un error, sino una huella de la estructura de los datos.

Soberanía cognitiva: capacidad de una comunidad o institución de tomar decisiones informadas sobre las tecnologías de IA que adopta, incluyendo la opción de correr modelos localmente.

Token: unidad mínima de información que procesa un modelo de lenguaje (palabra, parte de palabra, carácter).

Transformer: arquitectura de red neuronal basada en el mecanismo de atención, base de todos los LLM modernos.

Tree of Thoughts (ToT): evolución del CoT donde el modelo explora múltiples líneas de razonamiento en paralelo y selecciona la mejor.

Apéndice D – Preguntas para el aula (actividades)

D.1 ¿Dónde están mis datos?

Objetivo: conciencia sobre infraestructura y privacidad.
Procedimiento: los estudiantes investigan qué aplicaciones usan, dónde están sus servidores, qué dicen las políticas de privacidad. Se puede hacer con cualquier dispositivo o incluso con papel (buscando en políticas públicas).

D.2 Lo que la IA no sabe de acá

Objetivo: detectar sesgo geográfico y producir conocimiento local.
Procedimiento: los estudiantes preguntan a una IA sobre su propia localidad (historia, figuras destacadas, economía). Comparan con lo que ellos saben. Discuten por qué faltan cosas. Luego producen un texto (o audio, video) que corrija o amplíe la información faltante.

D.3 ¿Quién escribió estos datos?

Objetivo: comprender la cadena de trabajo detrás de la IA.
Procedimiento: seleccionar una respuesta de IA. Los estudiantes investigan: ¿qué

fuentes podría haber usado? ¿Qué perspectivas están representadas? ¿Cuáles están ausentes? ¿Quiénes pudieron haber anotado o limpiado los datos?

D.4 La IA como pregunta, no como respuesta

Objetivo: entrenar la formulación de preguntas.
Procedimiento: escribir un prompt que pida a la IA que genere preguntas en lugar de respuestas (ej. *"Generá 5 preguntas que un estudiante de secundaria debería hacerse sobre {tema}"*). Los estudiantes eligen la mejor pregunta y la responden ellos mismos, sin IA.

D.5 El error como oportunidad

Objetivo: pensamiento crítico sobre la IA.
Procedimiento: mostrar una respuesta incorrecta o sesgada de una IA (real o simulada). Los estudiantes identifican el problema, proponen por qué ocurrió (sesgo, alucinación, falta de contexto) y reformulan el prompt para corregirlo.

Hoja arrancable – Los 10 principios de la Inteligencia Relacional

(se puede fotocopiar o recortar para tener a mano)

1. **La IA calcula probabilidades, no piensa.** Usarla bien empieza por entender eso.
2. **El prompt es una configuración, no una conversación.** Cuanto más precisa, mejor.
3. **Los sesgos son estructurales.** No se eliminan, se compensan.
4. **Las alucinaciones son inevitables.** Verificar siempre.
5. **La soberanía no se pide. Se ejerce.** Preferir lo local cuando se pueda.
6. **Externalizá el producto. Preservá el proceso que desarrolla criterio.**
7. **La fricción es información.** Si la IA no entiende, preguntate qué no le dijiste.
8. **La responsabilidad no se delega.** La IA no rinde exámenes. Vos sí.
9. **La medida de una buena IA es si sirve en tu contexto real.**
10. **No hay destino. Hay práctica.** No se termina de aprender. Se vuelve a empezar.

Sierra Grande · Río Negro · Patagonia Argentina · 2026

Daniel Ricardo Novoa
Extensión de «Alfabetización Relacional: El lado humano de la inteligencia artificial»

Producido con herramientas de código abierto en ejecución local. Licencia Creative Commons BY-NC-SA 4.0.

Fin.